Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 22 (2020), Issue 6   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Viewpoints

Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers (e19284)
Guy Fagherazzi, Catherine Goetzinger, Mohammed Rashid, Gloria Aguayo, Laetitia Huiart. 11

Rethinking the Use of Mobile Apps for Dietary Assessment in Medical Research (e15619)
Wael Khazen, Jean-François Jeanne, Laëtitia Demaretz, Florent Schäfer, Guy Fagherazzi. 141

Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians (e15154)
Onur Asan, Alparslan Bayrak, Avishek Choudhury. 151

Artificial Intelligence and Big Data in Diabetes Care: A Position Statement of the Italian Association of Medical Diabetologists (e16922)
Nicoletta Musacchio, Annalisa Giancaterini, Giacomo Guaita, Alessandro Ozzello, Maria Pellegrini, Paola Ponzani, Giuseppina Russo, Rita Zilich, Alberto de Micheli. 158

The Adaptive Behavioral Components (ABC) Model for Planning Longitudinal Behavioral Technology-Based Health Interventions: A Theoretical Framework (e15563)
Sean Young. 169

Digital Health Equity and COVID-19: The Innovation Curve Cannot Reinforce the Social Gradient of Health (e19361)
Allison Crawford, Eva Serhal. 1310

Considerations for an Individual-Level Population Notification System for Pandemic Response: A Review and Prototype (e19930)
Mohammad Sakib, Zahid Butt, Plinio Morita, Mark Oremus, Geoffrey Fong, Peter Hall. 1335

Why African Americans Are a Potential Target for COVID-19 Infection in the United States (e19934)
Alireza Hamidian Jahromi, Anahid Hamidianjahromi. 1401

Leveraging User Experience to Improve Video Consultations in a Cardiology Practice During the COVID-19 Pandemic: Initial Insights (e19771)
Pieter Vandekerckhove, Yves Vandekerckhove, Rene Tavernier, Kelly De Jaegher, Marleen de Mul. 1492

Reviews

The Role of Virtual Reality in Improving Health Outcomes for Community-Dwelling Older Adults: Systematic Review (e17331)

Gordana Dermody, Lisa Whitehead, Graham Wilson, Courtney Glass. . . . 21

Therapeutic Alliance in Technology-Based Interventions for the Treatment of Depression: Systematic Review (e17195)

Eileen Wehmann, Moritz Köhnen, Martin Härter, Sarah Liebherz. . . . 38

Current Evidence for Continuous Vital Signs Monitoring by Wearable Wireless Devices in Hospitalized Adults: Systematic Review (e18636)

Jobbe Leenen, Crista Leerentveld, Joris van Dijk, Henderik van Westreenen, Lisette Schoonhoven, Gijsbert Patijn. . . . 72

Engaging Children and Young People in Digital Mental Health Interventions: Systematic Review of Modes of Delivery, Facilitators, and Barriers (e16317)

Shaun Liverpool, Catarina Mota, Célia Sales, Anja uš, Sara Carletto, Camellia Hancheva, Sónia Sousa, Sonia Cerón, Patricia Moreno-Peral, Giada Pietrabissa, Bettina Moltrecht, Randi Ulberg, Nuno Ferreira, Julian Edbrooke-Childs. . . . 94

Impact of the Method of Delivering Electronic Health Behavior Change Interventions in Survivors of Cancer on Engagement, Health Behaviors, and Health Outcomes: Systematic Review and Meta-Analysis (e16112)

Kate Furness, Mitchell Sarkies, Catherine Huggins, Daniel Croagh, Terry Haines. . . . 111

The Effectiveness of Digital Health Interventions in the Management of Musculoskeletal Conditions: Systematic Literature Review (e15617)

Stephanie Hewitt, Ruth Sephton, Gillian Yeowell. . . . 209

Behavior Change Techniques Included in Reports of Social Media Interventions for Promoting Health Behaviors in Adults: Content Analysis Within a Systematic Review (e16002)

Rosiane Simeon, Omar Dewidar, Jessica Trawin, Stephanie Duench, Heather Manson, Jordi Pardo Pardo, Jennifer Petkovic, Janet Hatcher Roberts, Peter Tugwell, Manosila Yoganathan, Justin Presseau, Vivian Welch. . . . 429

Patient Work and Their Contexts: Scoping Review (e16656)

Kathleen Yin, Joshua Jung, Enrico Coiera, Liliana Laranjo, Ann Blandford, Adeel Khoja, Wan-Tien Tai, Daniel Phillips, Annie Lau. . . . 501

Electronic Data Capture Versus Conventional Data Collection Methods in Clinical Pain Studies: Systematic Review and Meta-Analysis (e16480)

Lindsay Jibb, James Khan, Puneet Seth, Chitra Lalloo, Lauren Mulrooney, Kathryn Nicholson, Dominik Nowak, Harneel Kaur, Alyssandra Chee-A-Tow, Joel Foster, Jennifer Stinson. . . . 583

Health Equity in the Effectiveness of Web-Based Health Interventions for the Self-Care of People With Chronic Health Conditions: Systematic Review (e17849)

Sophie Turnbull, Christie Cabral, Alastair Hay, Patricia Lucas. . . . 662

Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review (e18301)

Alaa Abd-Alrazaq, Zeineb Safi, Mohannad Alajlani, Jim Warren, Mowafa Househ, Kerstin Denecke. . . . 1159

Original Papers

Two Decades of Research Using Taiwan’s National Health Insurance Claims Data: Bibliometric and Text Mining Analysis on PubMed (e18457)

Sheng-Feng Sung, Cheng-Yang Hsieh, Ya-Han Hu. . . . 53

A Digital Pornography Literacy Resource Co-Designed With Vulnerable Young People: Development of "The Gist" (e15964)

Angela Davis, Cassandra Wright, Stacey Murphy, Paul Dietze, Meredith Temple-Smith, Margaret Hellard, Megan Lim. 178

A Social Media–Based Support Group for Youth Living With HIV in Nigeria (SMART Connections): Randomized Controlled Trial (e18343)

Lisa Dulli, Kathleen Ridgeway, Catherine Packer, Kate Murray, Tolulope Mumuni, Kate Plourde, Mario Chen, Adesola Olumide, Oladosu Ojengbede, Donna McCarraher. 192

Efficacy of Three Low-Intensity, Internet-Based Psychological Interventions for the Treatment of Depression in Primary Care: Randomized Controlled Trial (e15845)

Margalida Gili, Adoración Castro, Azucena García-Palacios, Javier Garcia-Campayo, Fermin Mayoral-Cleries, Cristina Botella, Miquel Roca, Alberto Barceló-Soler, María Hurtado, MªTeresa Navarro, Amelia Villena, M Pérez-Ara, Pau Riera-Serra, Rosa Baños. 222

Evaluation of Digital Technologies Tailored to Support Young People's Self-Management of Musculoskeletal Pain: Mixed Methods Study (e18315)

Helen Slater, Jennifer Stinson, Joanne Jordan, Jason Chua, Ben Low, Chitra Lalloo, Quynh Pham, Joseph Cafazzo, Andrew Briggs. 242

Coach-Facilitated Web-Based Therapy Compared With Information About Web-Based Resources in Patients Referred to Secondary Mental Health Care for Depression: Randomized Controlled Trial (e15001)

Sarah MacLean, Daniel Corsi, Sadie Litchfield, Julia Kucharski, Kira Genise, Zeynep Selaman, Valerie Testa, Simon Hatcher. 264

User-Centered Design of a Mobile Health Intervention to Enhance Exacerbation-Related Self-Management in Patients With Chronic Obstructive Pulmonary Disease (Copilot): Mixed Methods Study (e15449)

Yvonne Korpershoek, Sander Hermsen, Lisette Schoonhoven, Marieke Schuurmans, Jaap Trappenburg. 283

Challenges and Benefits of an Internet-Based Intervention With a Peer Support Component for Older Adults With Depression: Qualitative Analysis of Textual Data (e17586)

Annie Chen, Krystal Slattery, Kathryn Tomasino, Caryn Rubanovich, Leland Bardsley, David Mohr. 302

Evaluation of an On-Demand Mental Health System for Depression Symptoms: Retrospective Observational Study (e17902)

Sarah Kunkle, Manny Yip, Watson Ξ, Justin Hunt. 314

A Blended Physiotherapy Intervention for Persons With Hemophilic Arthropathy: Development Study (e16631)

Merel Timmer, Corelien Kloek, Piet de Kleijn, Isolde Kuijlaars, Roger Schutgens, Cindy Veenhof, Martijn Pisters. 326

Young People's Experience of a Long-Term Social Media–Based Intervention for First-Episode Psychosis: Qualitative Analysis (e17570)

Lee Valentine, Carla McEnery, Shaunagh O'Sullivan, John Gleeson, Sarah Bendall, Mario Alvarez-Jimenez. 335

Evaluating Engagement in a Digital and Dietetic Intervention Promoting Healthy Weight Gain in Pregnancy: Mixed Methods Study (e17845)

Jane Willcox, Daniel Chai, Lawrence Beilin, Susan Prescott, Desiree Silva, Cliff Neppe, Rae-Chi Huang. 347

A Prospective Study of Usability and Workload of Electronic Medication Adherence Products by Older Adults, Caregivers, and Health Care Providers (e18073)

Tejal Patel, Jessica Ivo, Sadaf Faisal, Aidan McDougall, Jillian Carducci, Sarah Pritchard, Feng Chang. 358

Medical Mobile App Classification Using the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies: Interrater Reliability Study (e17457)

Khine Nwe, Mark Larsen, Natalie Nelissen, David Wong. 372

Reliability of Smartphone for Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores in Acute Ischemic Stroke Patients: Diagnostic Test Accuracy Study (e15893)
Kenichiro Sakai, Teppei Komatsu, Yasuyuki Iguchi, Hiroyuki Takao, Toshihiro Ishibashi, Yuichi Murayama. 380

Menstrual Cycle Length and Patterns in a Global Cohort of Women Using a Mobile Phone App: Retrospective Cohort Study (e17109)
Jessica Grieger, Robert Norman. 389

Stratification of Individual Symptoms of Contact Lens–Associated Dry Eye Using the iPhone App DryEyeRhythm: Crowdsourced Cross-Sectional Study (e18996)
Takenori Inomata, Masahiro Nakamura, Masao Iwagami, Akie Midorikawa-Inomata, Jaemyoung Sung, Keiichi Fujimoto, Yuichi Okumura, Atsuko Eguchi, Nanami Iwata, Maria Miura, Kenta Fujio, Ken Nagino, Satoshi Hori, Kazuo Tsubota, Reza Dana, Akira Murakami. 404

Challenges and Successes in Raising a Child With Type 1 Diabetes and Autism Spectrum Disorder: Mixed Methods Study (e17184)
Tamara Oser, Sean Oser, Jessica Parascando, Lee Grisolano, Kanthi Krishna, Daniel Hale, Michelle Litchman, Shideh Majidi, Paul Haidet. . .
4 1 7

A Social Media Study on the Associations of Flavored Electronic Cigarettes With Health Symptoms: Observational Study (e17496)
Long Chen, Xinyi Lu, Jianbo Yuan, Joyce Luo, Jiebo Luo, Zidian Xie, Dongmei Li. 445

Association Between HIV-Related Tweets and HIV Incidence in the United States: Infodemiology Study (e17196)
Robin Stevens, Stephen Bonett, Jacqueline Bannon, Deepti Chittamuru, Barry Slaff, Safa Browne, Sarah Huang, José Bauermeister. 456

User Perceptions of Different Electronic Cigarette Flavors on Social Media: Observational Study (e17280)
Xinyi Lu, Long Chen, Jianbo Yuan, Joyce Luo, Jiebo Luo, Zidian Xie, Dongmei Li. 466

Exploring Public Awareness of Overwork Prevention With Big Data From Google Trends: Retrospective Analysis (e18181)
Ro-Ting Lin, Yawen Cheng, Yan-Cheng Jiang. 477

Impact of Age on Patients’ Communication and Technology Preferences in the Era of Meaningful Use: Mixed Methods Study (e13470)
Martina Clarke, Ann Fruhling, Marilyn Sitorius, Thomas Windle, Tamara Bernard, John Windle. 487

Physiotherapy With Telerehabilitation in Patients With Complicated Postoperative Recovery After Esophageal Cancer Surgery: Feasibility Study (e16056)
Maarten van Egmond, Raoul Engelbert, Jean Klinkenbijl, Mark van Berge Henegouwen, Marike van der Schaaf. 516

Evaluating Safety and Efficacy of Follow-up for Patients With Abdominal Pain Using Video Consultation (SAVED Study): Randomized Controlled Trial (e17417)
Dinesh Gunasekeran, Zhenghong Liu, Win Tan, Joshua Koh, Chiu Cheong, Lay Tan, Chee Lau, Gaik Phuah, Newsie Manuel, Che Chia, Gek Seng, Nancy Tong, May Huin, Suzette Dulce, Susan Yap, Kishanti Ponampalam, Hao Ying, Marcus Ong, R Ponampalam. 529

Communication Strategies Used to Obtain Clinical Histories Before Remotely Prescribing Antibiotics for Postal Treatment of Uncomplicated Genital Chlamydia: Service Evaluation (e15970)
Hannah McCulloch, Jonathan Syred, Gillian Holdsworth, Chris Howroyd, Elena Ardines, Paula Baraitser. 539

Perspectives of Psychotherapists and Psychiatrists on Mental Health Care Integration Within Primary Care Via Video Consultations: Qualitative Preimplementation Study (e17569)
Mariell Hoffmann, Michel Wensing, Frank Peters-Klimm, Joachim Szecsenyi, Mechthild Hartmann, Hans-Christoph Friederich, Markus Haun.
5 5 1

Improvements in Patient Monitoring in the Intensive Care Unit: Survey Study (e19091)
Akira-Sebastian Poncette, Lina Mosch, Claudia Spies, Malte Schmieding, Fridtjof Schiefenhövel, Henning Krampe, Felix Balzer. . . . 562

Active Surveillance of Adverse Events Following Human Papillomavirus Vaccination: Feasibility Pilot Study Based on the Regional Health Care Information Platform in the City of Ningbo, China (e17446)
Zhike Liu, Liang Zhang, Yu Yang, Ruogu Meng, Ting Fang, Ying Dong, Ning Li, Guozhang Xu, Siyan Zhan. . . . 573

Attitudes and Use of Information and Communication Technologies in Older Adults With Mild Cognitive Impairment or Early Stages of Dementia and Their Caregivers: Cross-Sectional Study (e17253)
Jose Guzman-Parra, Pilar Barnestein-Fonseca, Gloria Guerrero-Pertiñez, Peter Anderberg, Luis Jimenez-Fernandez, Esperanza Valero-Moreno, Jessica Goodman-Casanova, Antonio Cuesta-Vargas, Maite Garolera, Maria Quintana, Rebeca García-Betances, Evi Lemmens, Johan Sanmartin Berglund, Fermin Mayoral-Cleries. . . . 616

Gender and the Digital Divide Across Urban Slums of New Delhi, India: Cross-Sectional Study (e14714)
Ashish Joshi, Bhavya Malhotra, Chioma Amadi, Menka Loomba, Archa Misra, Shruti Sharma, Arushi Arora, Jaya Amatya. . . . 631

Internet and Health Information Technology Use and Psychological Distress Among Older Adults With Self-Reported Vision Impairment: Case-Control Study (e17294)
Namkee Choi, Diana DiNitto, Othelia Lee, Bryan Choi. . . . 648

Lifestyle Segmentation to Explain the Online Health Information–Seeking Behavior of Older Adults: Representative Telephone Survey (e15099)
Winja Weber, Anne Reinhardt, Constanze Rossmann. . . . 683

Views of Swedish Elder Care Personnel on Ongoing Digital Transformation: Cross-Sectional Study (e15450)
Katarina Baudin, Christine Gustafsson, Susanne Frennert. . . . 697

The Demographic Representativeness and Health Outcomes of Digital Health Station Users: Longitudinal Study (e14977)
Leah Flitcroft, Won Chen, Denny Meyer. . . . 707

Continuous Monitoring of Vital Signs in the General Ward Using Wearable Devices: Randomized Controlled Trial (e15471)
Mariska Weenk, Sebastian Bredie, Mats Koeneman, Gijs Hesselink, Harry van Goor, Tom van de Belt. . . . 722

Internet Influence of Assisted Reproduction Technology Centers in China: Qualitative Study Based on WeChat Official Accounts (e17997)
Fang Shao, Zhiqiang He, Zheng Zhu, Xiang Wang, Jianping Zhang, Jinhua Shan, Jiajia Pan, Hui Wang. . . . 733

Enhancing Patient Experience With Internet Protocol Addressable Digital Light-Emitting Diode Lighting in Imaging Environments: A Phase I Study (e11839)
Melanie Knopp, Katherine Binzel, Chadwick Wright, Jun Zhang, Michael Knopp. . . . 747

Metrics for Outpatient Portal Use Based on Log File Analysis: Algorithm Development (e16849)
Gennaro Di Tosto, Ann McAlearney, Naleef Fareed, Timothy Huerta. . . . 756

The Impact of Digital-First Consultations on Workload in General Practice: Modeling Study (e18203)
Chris Salisbury, Mairead Murphy, Polly Duncan. . . . 773

Applying Machine Learning to Daily-Life Data From the TrackYourTinnitus Mobile Health Crowdsensing Platform to Predict the Mobile Operating System Used With High Accuracy: Longitudinal Observational Study (e15547)
Rüdiger Pryss, Winfried Schlee, Burkhard Hoppenstedt, Manfred Reichert, Myra Spiliopoulou, Berthold Langguth, Marius Breitmayer, Thomas Probst. . . . 784

Modifications to Electronic Nicotine Delivery Systems: Content Analysis of YouTube Videos (e17104)

Zachary Massey, Yachao Li, Jessica Holli, Victoria Churchill, Bo Yang, Katherine Henderson, David Ashley, Jidong Huang, Lucy Popova. . . .
7 9 9

The Use of Mobile Personal Health Records for Hemoglobin A1c Regulation in Patients With Diabetes: Retrospective Observational Study (e15372)

Dongjin Seo, Yu Park, Yura Lee, Ji Kim, Joong-Yeol Park, Jae-Ho Lee. . . . 809

Comparisons Between Hypothesis- and Data-Driven Approaches for Multimorbidity Frailty Index: A Machine Learning Approach (e16213)

Li-Ning Peng, Fei-Yuan Hsiao, Wei-Ju Lee, Shih-Tsung Huang, Liang-Kung Chen. . . . 823

Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial (e16772)

Windy Chan, Angela Leung. . . . 834

Teaching the National Institutes of Health Stroke Scale to Paramedics (E-Learning vs Video): Randomized Controlled Trial (e18358)

Avinash Koka, Laurent Suppan, Philippe Cottet, Emmanuel Carrera, Loric Stuby, Mélanie Suppan. . . . 850

The Model of Gamification Principles for Digital Health Interventions: Evaluation of Validity and Potential Utility (e16506)

Mark Floryan, Philip Chow, Stephen Schueller, Lee Ritterband. . . . 858

Using Geocaching to Promote Active Aging: Qualitative Study (e15339)

Silvia Fornasini, Marco Dianti, Alessandro Bacchiega, Stefano Forti, Diego Conforti. . . . 869

Understanding Weight Loss via Online Discussions: Content Analysis of Reddit Posts Using Topic Modeling and Word Clustering Techniques (e13745)

Yang Liu, Zhijun Yin. . . . 882

Superusers' Engagement in Asthma Online Communities: Asynchronous Web-Based Interview Study (e18185)

Anna De Simoni, Anjali Shah, Olivia Fulton, Jasmine Parkinson, Aziz Sheikh, Pietro Panzarasa, Claudia Pagliari, Neil Coulson, Chris Griffiths.
8 9 6

Examining Social Capital, Social Support, and Language Use in an Online Depression Forum: Social Network and Content Analysis (e17365)

Wenjing Pan, Bo Feng, Cuihua Shen. . . . 911

Usability of a Consumer Health Informatics Tool Following Completion of a Clinical Trial: Focus Group Study (e17708)

Hwayoung Cho, Tiffany Porras, Gabriella Flynn, Rebecca Schnall. . . . 925

Facebook Recruitment Using Zip Codes to Improve Diversity in Health Research: Longitudinal Observational Study (e17554)

Cornelia Pechmann, Connor Phillips, Douglas Calder, Judith Prochaska. . . . 936

Spelling Errors and Shouting Capitalization Lead to Additive Penalties to Trustworthiness of Online Health Information: Randomized Experiment With Laypersons (e15171)

Harry Witchel, Georgina Thompson, Christopher Jones, Carina Westling, Juan Romero, Alessia Nicotra, Bruno Maag, Hugo Critchley. . . . 957

Web-Based Health Information Following the Renewal of the Cervical Screening Program in Australia: Evaluation of Readability, Understandability, and Credibility (e16701)
Olivia Mac, Amy Thayre, Shumei Tan, Rachael Dodd. 971

Cigarette Packs With URLs Leading to Tobacco Company Websites: Content Analysis (e15160)
Caitlin Weiger, Katherine Smith, Amy Hong, Joanna Cohen. 985

Privacy-Preserving Record Linkage of Deidentified Records Within a Public Health Surveillance System: Evaluation Study (e16757)
Long Nguyen, Mark Stoové, Douglas Boyle, Denton Callander, Hamish McManus, Jason Asselin, Rebecca Guy, Basil Donovan, Margaret Hellard, Carol El-Hayek. 997

Association of eHealth Literacy With Lifestyle Behaviors in University Students: Questionnaire-Based Cross-Sectional Study (e18155)
Saki Tsukahara, Satoshi Yamaguchi, Futaba Igarashi, Reiko Uruma, Naomi Ikuina, Kaori Iwakura, Keisuke Koizumi, Yasunori Sato. 1009

Disseminating the Foundations of Knowledge Translation and Patient Engagement Science Through the KnowledgeNudge Blog and Twitter Profile: Quantitative Descriptive Evaluation (e15351)
Kathryn Sibley, Masood Khan, Patricia Roche, Patrick Faucher, Carly Leggett. 1020

Prediction of (Non)Participation of Older People in Digital Health Research: Exergame Intervention Study (e17884)
Arianna Poli, Susanne Kelfve, Leonie Klompstra, Anna Strömberg, Tiny Jaarsma, Andreas Motel-Klingebiel. 1029

Physiological State and Learning Ability of Students in Normal and Virtual Reality Conditions: Complexity-Based Analysis (e17945)
Mohammad Babini, Vladimir Kulish, Hamidreza Namazi. 1041

Simulator Pre-Screening of Underprepared Drivers Prior to Licensing On-Road Examination: Clustering of Virtual Driving Test Time Series Data (e13995)
David Grethlein, Flaura Winston, Elizabeth Walshe, Sean Tanner, Venk Kandadai, Santiago Ontañón. 1050

Using Virtual Reality to Improve Apathy in Residential Aged Care: Mixed Methods Study (e17632)
Dimitrios Saredakis, Hannah Keage, Megan Corlis, Tobias Loetscher. 1064

Internet Use, eHealth Literacy, and Dietary Supplement Use Among Young Adults in Pakistan: Cross-Sectional Study (e17014)
Amina Tariq, Shanchita Khan, Amna Basharat. 1086

Factors Associated With the Actual Behavior and Intention of Rating Physicians on Physician Rating Websites: Cross-Sectional Study (e14417)
Xi Han, Bei Li, Tingting Zhang, Jiabin Qu. 1095

Associations Between Commercial App Use and Physical Activity: Cross-Sectional Study (e17152)
Jasmine Petersen, Eva Kemps, Lucy Lewis, Ivanka Prichard. 1109

Effect of BMI and Perceived Importance of Health on the Health Behavior of College Students: Cross-Sectional Study (e17640)
Wan-Chen Hsu, Chia-Hsun Chiang. 1122

Developing and Applying a Formative Evaluation Framework for Health Information Technology Implementations: Qualitative Investigation (e15068)
Kathrin Cresswell, Robin Williams, Aziz Sheikh. 1128

Patient-Centric Scheduling With the Implementation of Health Information Technology to Improve the Patient Experience and Access to Care: Retrospective Case-Control Analysis (e16451)
Sukyung Chung, Meghan Martinez, Dominick Frosch, Veena Jones, Albert Chan. 1138

Effect of Speech Recognition on Problem Solving and Recall in Consumer Digital Health Tasks: Controlled Laboratory Experiment (e14827)
Jessica Chen, David Lyell, Liliana Laranjo, Farah Magrabi. 1149

Using Blockchain Technology to Mitigate Challenges in Service Access for the Homeless and Data Exchange Between Providers: Qualitative Study (e16887)
Anjum Khurshid, Vivian Rajeswaren, Steven Andrews. 1174

An Architecture and Management Platform for Blockchain-Based Personal Health Record Exchange: Development and Usability Study (e16748)
Hsiu-An Lee, Hsin-Hua Kung, Jai Udayasankaran, Boonchai Kijsanayotin, Alvin B Marcelo, Louis Chao, Chien-Yeh Hsu. 1185

Inferring Sexually Transmitted Infection Risk From Attractiveness in Online Dating Among Adolescents and Young Adults: Exploratory Study (e14242)
Tamar Krishnamurti, Alexander Davis, Baruch Fischhoff. 1200

Toward the Development of Data Governance Standards for Using Clinical Free-Text Data in Health Research: Position Paper (e16760)
Kerina Jones, Elizabeth Ford, Nathan Lea, Lucy Griffiths, Lamiece Hassan, Sharon Heys, Emma Squires, Goran Nenadic. 1209

Healthcare Research and Analytics Data Infrastructure Solution: A Data Warehouse for Health Services Research (e18579)
Bunyamin Ozaydin, Ferhat Zengul, Nurettin Oner, Sue Feldman. 1225

Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach (e18585)
Cheng-Sheng Yu, Yu-Jiun Lin, Chang-Hsien Lin, Shiyng-Yu Lin, Jenny Wu, Shy-Shin Chang. 1241

Patients' Perceptions of Barriers and Facilitators to the Adoption of E-Hospitals: Cross-Sectional Study in Western China (e17221)
Peiyi Li, Yunmei Luo, Xuexin Yu, Jin Wen, Elizabeth Mason, Weimin Li, Mohammad Jalali. 1259

An Intervention Delivered by Mobile Phone Instant Messaging to Increase Acceptability and Use of Effective Contraception Among Young Women in Bolivia: Randomized Controlled Trial (e14073)
Ona McCarthy, Carolina Aliaga, Maria Torrico Palacios, Jhonny López Gallardo, Silvia Huaynoca, Baptiste Leurent, Phil Edwards, Melissa Palmer, Irrfan Ahamed, Caroline Free. 1276

Data Validation and Verification Using Blockchain in a Clinical Trial for Breast Cancer: Regulatory Sandbox (e18938)
Tomonobu Hirano, Tomomitsu Motohashi, Kosuke Okumura, Kentaro Takajo, Taiyo Kuroki, Daisuke Ichikawa, Yutaka Matsuoka, Eisuke Ochi, Taro Ueno. 1292

Online Information Exchange and Anxiety Spread in the Early Stage of the Novel Coronavirus (COVID-19) Outbreak in South Korea: Structural Topic Model and Network Analysis (e19455)
Wonkwang Jo, Jaeho Lee, Junli Park, Yeol Kim. 1315

COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan (e20021)
Peng-Wei Wang, Wei-Hsin Lu, Nai-Ying Ko, Yi-Lung Chen, Dian-Jeng Li, Yu-Ping Chang, Cheng-Fang Yen. 1348

Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States (e19264)
Malcolm Fisk, Anne Livingstone, Sabrina Pit. 1356

Development of an openEHR Template for COVID-19 Based on Clinical Guidelines (e20239)
Mengyang Li, Heather Leslie, Bin Qi, Shan Nan, Hongshuo Feng, Hailing Cai, Xudong Lu, Huilong Duan. 1369

Epidemiological Observations on the Association Between Anosmia and COVID-19 Infection: Analysis of Data From a Self-Assessment Web Application (e19855)
Fabrice Denis, Simon Galmiche, Aurélien Dinh, Arnaud Fontanet, Arnaud Scherpereel, Francois Benezit, François-Xavier Lescure. 1384

Information Technology–Based Management of Clinically Healthy COVID-19 Patients: Lessons From a Living and Treatment Support Center Operated by Seoul National University Hospital (e19938)
Ye Bae, Kyung Kim, Sae Choi, Taehoon Ko, Chang Jeong, BeLong Cho, Min Kim, EunKyo Kang. 1390

Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo (e19981)
Zhuo-Ying Tao, Guang Chu, Colman McGrath, Fang Hua, Yiu Leung, Wei-Fa Yang, Yu-Xiong Su. 1405

Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States (e19782)
Siyue Li, Bo Feng, Wang Liao, Wenjing Pan. 1419

Clarification of Misleading Perceptions of COVID-19 Fatality and Testing Rates in Italy: Data Analysis (e19825)
Davide Tosi, Alessandro Verde, Manuela Verde. 1431

Human Factor Considerations in Using Personal Protective Equipment in the COVID-19 Pandemic Context: Binational Survey Study (e19947)
Avi Parush, Oren Wacht, Ricardo Gomes, Amit Frenkel. 1438

Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study (e20185)
Jeremy Huckins, Alex daSilva, Weichen Wang, Elin Hedlund, Courtney Rogers, Subigya Nepal, Jialing Wu, Mikio Obuchi, Eilis Murphy, Meghan Meyer, Dylan Wagner, Paul Holtzheimer, Andrew Campbell. 1450

A Double Triage and Telemedicine Protocol to Optimize Infection Control in an Emergency Department in Taiwan During the COVID-19 Pandemic: Retrospective Feasibility Study (e20586)
Chien-Hao Lin, Wen-Pin Tseng, Jhong-Lin Wu, Joyce Tay, Ming-Tai Cheng, Hooi-Nee Ong, Hao-Yang Lin, Yi-Ying Chen, Chih-Hsien Wu, Jiun-Wei Chen, Shey-Ying Chen, Chang-Chuan Chan, Chien-Hua Huang, Shyr-Chyr Chen. 1469

Virtual Management of Patients With Cancer During the COVID-19 Pandemic: Web-Based Questionnaire Study (e19691)
Emad Tashkandi, Ahmed Zeeneldin, Amal AlAbdulwahab, Omima Elemam, Shereef Elsamany, Wasil Jastaniah, Shaker Abdullah, Mohammad Alfayez, Abdul Jazieh, Humaid Al-Shamsi. 1481

Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation (e19659)
Viroj Tangcharoensathien, Neville Calleja, Tim Nguyen, Tina Purnat, Marcelo D'Agostino, Sebastian Garcia-Saiso, Mark Landry, Arash Rashidian, Clayton Hamilton, Abdelhalim AbdAllah, Ioana Ghiga, Alexandra Hill, Daniel Hougendobler, Judith van Andel, Mark Nunn, Ian Brooks, Pier Sacco, Manlio De Domenico, Philip Mai, Anatoliy Gruzd, Alexandre Alaphilippe, Sylvie Briand. 1498

COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep Learning Framework With a Single Chest CT Image: Model Development and Validation (e19569)
Hoon Ko, Heewon Chung, Wu Kang, Kyung Kim, Youngbin Shin, Seung Kang, Jae Lee, Young Kim, Nan Kim, Hyunseok Jung, Jinseok Lee. .
1 5 0 6

A COVID-19 Risk Assessment Decision Support System for General Practitioners: Design and Development Study (e19786)
Ying Liu, Zhixiao Wang, Jingjing Ren, Yu Tian, Min Zhou, Tianshu Zhou, Kangli Ye, Yinghao Zhao, Yunqing Qiu, Jingsong Li. 1519

Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province (e19636)
Jitian Li, Zhe Chen, Yifei Nie, Yan Ma, Qiaoyun Guo, Xiaofeng Dai. 1534

Perceptions About Technologies That Help Community-Dwelling Older Adults Remain at Home: Qualitative Study (e17930)
Henk Verloo, Thomas Kampel, Nicole Vidal, Filipa Pereira. 1545

Commentary

How to Fight an Infodemic: The Four Pillars of Infodemic Management (e21820)
Gunther Eysenbach. 135

Short Papers

Evaluating the Need to Address Digital Literacy Among Hospitalized Patients: Cross-Sectional Observational Study (e17519)
Hanna Vollbrecht, Vineet Arora, Sebastian Otero, Kyle Carey, David Meltzer, Valerie Press. 1079

Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis (e19787)
Jean Pépin, Rosa Bruno, Rui-Yi Yang, Vincent Vercamer, Paul Jouhaud, Pierre Escourrou, Pierre Boutouyrie. 1463

Corrigenda and Addendas

Correction: Considerations for an Individual-Level Population Notification System for Pandemic Response: A Review and Prototype (e21634)
Mohammad Sakib, Zahid Butt, Plinio Morita, Mark Oremus, Geoffrey Fong, Peter Hall. 1306

Correction: Offline Digital Education for Postregistration Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration (e20316)
Pawel Posadzki, Malgorzata Bala, Bhone Kyaw, Monika Semwal, Ushashree Divakar, Magdalena Koperny, Agnieszka Sliwka, Josip Car. . .
1 3 0 8

Viewpoint

Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers

Guy Fagherazzi[1], MSc, PhD; Catherine Goetzinger[1], MSc; Mohammed Ally Rashid[1], MD; Gloria A Aguayo[1], MD, PhD; Laetitia Huiart[1], MD, PhD

[1] Luxembourg Institute of Health, Strassen, Luxembourg

Corresponding Author:
Guy Fagherazzi, MSc, PhD
Luxembourg Institute of Health
1 A-B Rue Thomas Edison
Strassen, 1445
Luxembourg
Phone: 352 33669396334
Email: guy.fagherazzi@lih.lu

Abstract

The coronavirus disease (COVID-19) pandemic has created an urgent need for coordinated mechanisms to respond to the outbreak across health sectors, and digital health solutions have been identified as promising approaches to address this challenge. This editorial discusses the current situation regarding digital health solutions to fight COVID-19 as well as the challenges and ethical hurdles to broad and long-term implementation of these solutions. To decrease the risk of infection, telemedicine has been used as a successful health care model in both emergency and primary care. Official communication plans should promote facile and diverse channels to inform people about the pandemic and to avoid rumors and reduce threats to public health. Social media platforms such as Twitter and Google Trends analyses are highly beneficial to model pandemic trends as well as to monitor the evolution of patients' symptoms or public reaction to the pandemic over time. However, acceptability of digital solutions may face challenges due to potential conflicts with users' cultural, moral, and religious backgrounds. Digital tools can provide collective public health benefits; however, they may be intrusive and can erode individual freedoms or leave vulnerable populations behind. The COVID-19 pandemic has demonstrated the strong potential of various digital health solutions that have been tested during the crisis. More concerted measures should be implemented to ensure that future digital health initiatives will have a greater impact on the epidemic and meet the most strategic needs to ease the life of people who are at the forefront of the crisis.

(J Med Internet Res 2020;22(6):e19284) doi:10.2196/19284

KEYWORDS

coronavirus; COVID-19; digital health; eHealth; digital technology; health care; surveillance; communication; review; epidemiology; infodemiology; public health

Background

Countries around the world have been affected by the COVID-19 pandemic since December 2019 [1], and the health care systems in these countries are rapidly adapting to the increasing demand. The World Health Organization (WHO) has called for coordinated mechanisms to support the response to the outbreak across health sectors, and digital health solutions have been identified as one of the most promising approaches to address this challenge in modern societies [2].

The COVID-19 pandemic is singular in many ways. First, in terms of number of people infected, transmissibility, and spectrum of clinical severity, it has had a greater impact to date than previous epidemics such as pandemic influenza, Middle East respiratory syndrome (MERS), severe acute respiratory syndrome (SARS), or Ebola virus [3]. Second, COVID-19 can be considered as the first true global epidemic of this magnitude in the digital era; digital health solutions, which have reached a certain level of maturity but are not widely deployed and accepted yet, can play a major part in our response to the crisis [4]. Indeed, the COVID-19 pandemic is occurring in an era of massive technological advancement. Digital tools can effectively support institutions during a pandemic by facilitating the immediate widespread distribution of information [5], tracking transmission in real time, creating virtual venues for meetings or day-to-day operations, and providing telemedicine visits for patients [6]. However, some digital health strategies and tools may face challenges associated with barriers to access, acceptability, and ethical issues. For instance, some governments worldwide are responding to this public health emergency with an unprecedented array of surveillance tools designed to identify and track people who may be infectious [7]. These measures have escalated existing debates regarding individual privacy and government oversight of private citizens.

The COVID-19 pandemic is providing a starting point to discuss how digital health solutions can and should be leveraged to address this unprecedented crisis [8]. This editorial offers an overview of the digital health strategies that have been employed worldwide to fight COVID-19 to date as well as a discussion of the challenges to developing more meaningful [9] and ethical solutions in the near future.

Digital Health Care Models

New health care models are needed during the COVID-19 pandemic. Due to the high transmission rate of COVID-19, most countries have ordered strict lockdowns; hence, patient-physician communication and visits are challenging. In the era of digital health technologies, the focus on new models has shifted to telehealth (virtual visits, virtual care), mobile apps (remote patient monitoring), and websites and chatbots (risk assessment, screening, triage). Figure 1 describes the main stakeholders at the forefront of the COVID_19 crisis (the population, the health care systems, and the research and health technology environment in coordination with public health agencies or governments) with significant illustrations of efficient digital health strategies that have been implemented to date at various levels.

Figure 1. The COVID-19 digital health ecosystem. AI: artifical intelligence; e-prescribing: electronic prescribing; tech: technology.

Telemedicine and Remote Patient Monitoring Systems

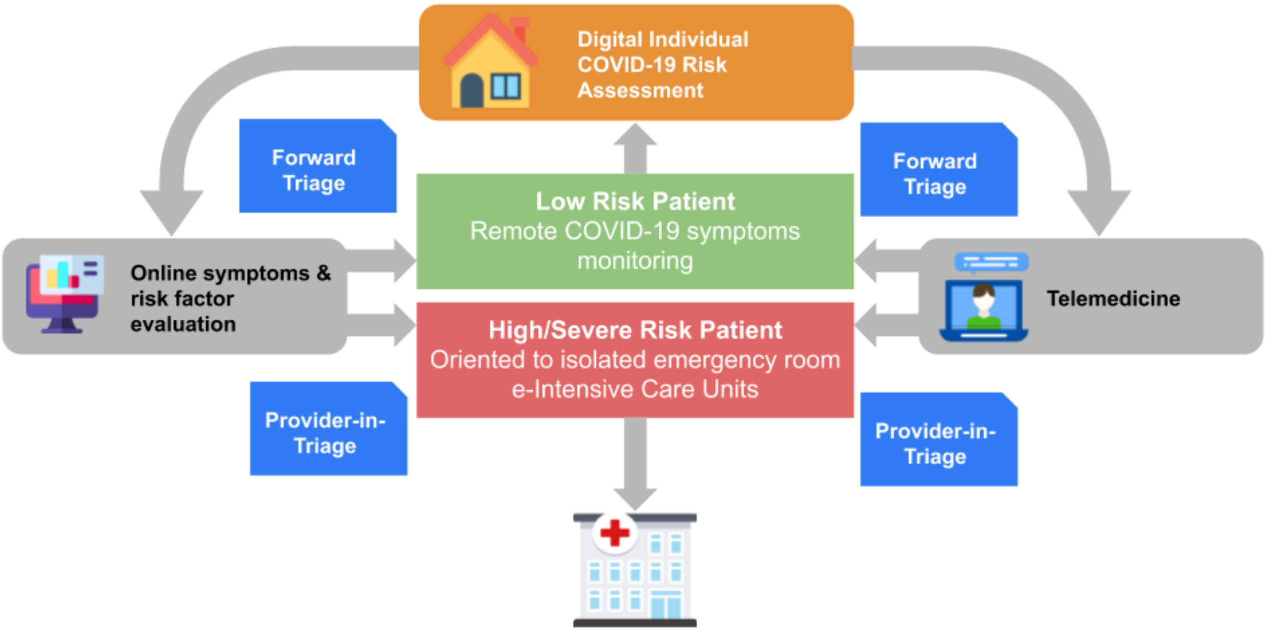
Telemedicine allows patients to receive care at home [10], thus avoiding the spread of COVID-19 in overcrowded emergency or waiting rooms [11]. Telemedicine not only supports secured care for COVID-19 patients but also allows routine primary care and electronic prescriptions (e-prescriptions). In addition, health care professionals can screen and monitor symptoms in real time, provide useful medical advice when needed, and keep stable patients at home away from overloaded hospitals [12].

Additionally, these tools can serve as useful resources to rapidly collect meaningful information on large cohorts of patients to study the evolution of their symptoms in real time; this increases understanding of the different clinical phenotypes of people infected with COVID-19 and enables the study of its long-term health consequences. Such a prospective cohort was recently established in Luxembourg (Predi-COVID) to study factors associated with COVID-19 disease severity; this study uses data from the national telesurveillance system that is used for virtually all patients who tested positive for COVID-19 and combines the data with biological sampling, electronic patient-reported outcomes, and innovative digital data collections such as smartphone-based voice recording to identify vocal biomarkers of respiratory syndromes. In another study, Wang et al [13] described armband sensors that can remotely measure the accuracy of hand hygiene based on WHO procedures.

Triage and Risk Assessment

The COVID-19 virus has caused panic and uncertainty within populations; to ensure better allocation of resources, telemedicine based on videoconferences can be used not only to comfort patients but also for triage [14]. Hollander et al [15] highlighted that telemedicine can serve as a triage system in different phases. The first stage is “forward triage;” here, patients are categorized as possibly COVID-19 infected or uninfected before they arrive at the hospital. In addition, telemedicine ensures close monitoring of less severely infected patients at home by allowing communication around the clock. The second phase of triage occurs in the hospital and is referred to as the “provider-in-triage model.” In this phase, patients at high risk for COVID-19 are screened by rapid evaluation of clinical characteristics and testing. Patients are kept in isolated emergency rooms, where they are given tablet devices to communicate with professionals [15]; the tablets are cleaned between patients. To reduce risk of infection among health care professionals, hospitals have also implemented electronic intensive care unit monitoring. For example, a few countries have launched within-hospital telemedicine programs. A robot enters the room of a patient infected with COVID-19 to monitor both vital and cardiac signs and enable communication between the patient and health care professionals where applicable. All the robot’s actions are controlled by nurses or physicians outside the room [16]. Figure 2 shows the typical path of an individual from the time they have questions about their health or symptoms potentially related to COVID-19. Digital technologies now enable online symptom checks, video consultations, and e-prescriptions and allow forward triage instead of overloading hospitals and increasing the spread of the virus. Low-risk patients can be regularly monitored from home, whereas patients with more severe symptoms can be monitored at hospitals with a provider-in-triage using digital tools.

Figure 2. Risk assessment and triage of COVID-19 in a digital health care model. COVID-19: coronavirus disease.

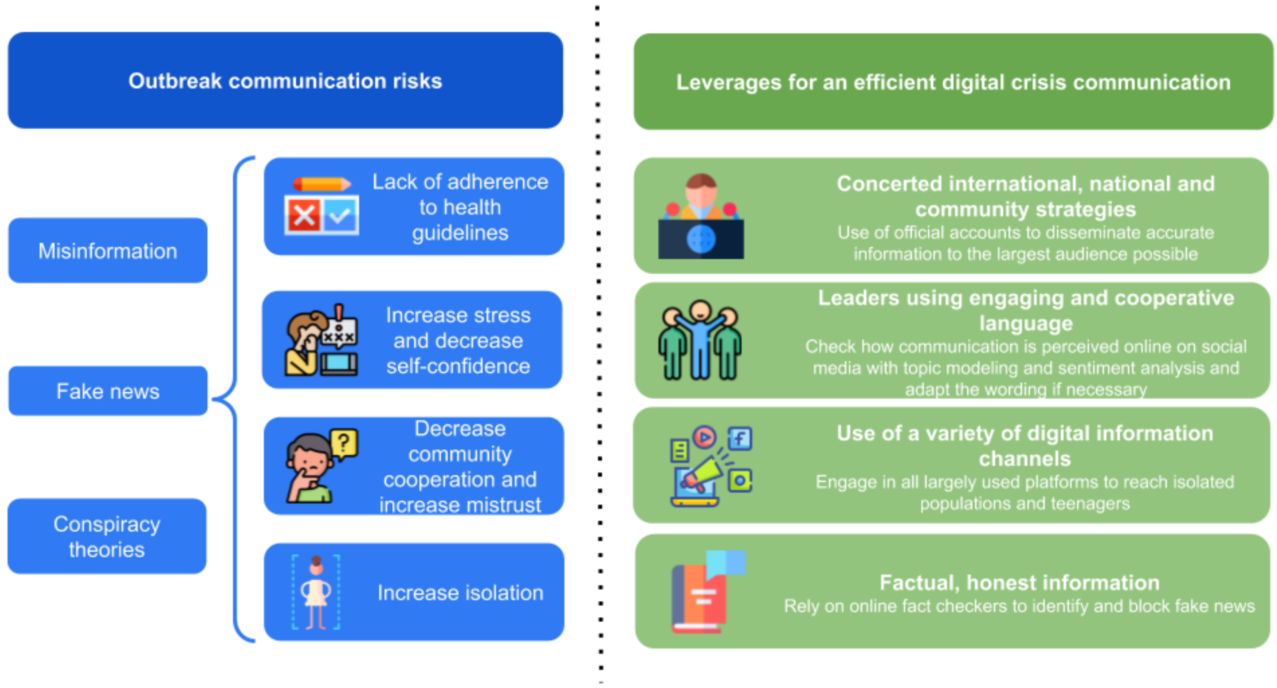


Web-based symptom trackers such as maladiecoronavirus.fr [17] in France and beatcovid19now.org [18] in Australia were rapidly established to help individuals assess their risk for COVID-19; these trackers are being updated as knowledge of COVID-19 evolves. Site users must answer several questions to feed the risk prediction algorithms; then, based on risk stratification, the website advises the user to call a hotline to receive further instructions or to stay at home and follow international guidelines. Similarly, chatbots such as Symptoma are systems that digitally assist individuals in performing risk assessment of COVID-19 [19].

Digital Communication During the Pandemic

Communication between political leaders and scientific authorities is crucial during a pandemic crisis [20] (Figure 3). Since the start of the COVID-19 pandemic, disinformation, “fake news,” and conspiracy theories about the origin, spread, and treatment of the disease have been present on the internet and in government communications, social media, and SMS text messages. These communication problems also cause confusion and mistrust; the consequences are negative at both the individual and community levels. To provide the most useful information in a timely fashion, effective digital risk communication has been established internationally by WHO as well as nationally and regionally through an active factual information campaign using several digital channels. Good communication strategies help manage people's fear and increase the likelihood that they will adhere to the measures imposed on them during the crisis. Risk communication promotes community engagement, avoids or decreases the profusion of rumors, and reduces threats to public health. However, poor risk communication can create a great deal of confusion, frustration, and future mistrust of leaders. Communication in epidemics must not only be clear and unambiguous but must also reach the target population quickly (health workers, local authorities, and the public). For instance, guidelines around practicing social distancing have been judged to be unclear and prone to misunderstanding. These guidelines should be followed with more direct messages, such as “stay at home,” especially to clarify the concept of social distancing for vulnerable populations or younger people. The WHO webpage uses the phrase “physical distance” in addition to “social distance” to emphasize this necessity. However, social contact must be stimulated, especially for older persons [21]. Digital methods appear to be promising solutions, as they enable people to maintain physical distance while keeping informed or maintaining contact with others; for example, the WhatsApp messaging platform provides health alerts in different languages [22].

Figure 3. Communication risks and strategies during disease outbreaks.



In addition, digital health can play a major role at a governmental level. Government initiatives have been taken to provide guidelines and information on government websites to inform the population about appropriate behavior during the lockdown period, hand hygiene, etc. Most of these initiatives were supported by international guidelines from WHO or the US Centers for Disease Control and Prevention (CDC).

The use of apps or instant messaging services to enable ministries or health agencies to contact health workers remotely and to inform government officials about the number of cases, symptoms, and prevention measures are just a few examples of measures that have been efficiently used in previous pandemics [23] or early in the COVID-19 crisis [22]. A crucial consequence of a bad risk communication strategy is the likelihood of increased levels of anxiety within the population. Contradictory information and overexposure to sensationalist headlines from mass media and social media can affect people’s mental health and increase the general level of fear. Lockdown can amplify the problem due to increased exposure to social media and rumors as well as lack of confrontation of ideas and debate in the real world [24]. Consequently, lockdown reinforces the importance of institutional development of digital strategies. Therefore, during the outbreak, official communication plans should promote easily accessible and diverse channels of digital communication at all stages and for different purposes (global, regional, and community-based communication). These channels should be integrated as an information system, as a communication tool, and for detecting and supporting people at risk [25]. An example of a communication strategy that was developed very quickly and effectively to fight fear and rumors among a population is the sending of messages by the government of Singapore through WhatsApp [26]. Finally, a strategy to include marginalized and vulnerable populations should be planned. Such a strategy should involve establishing multiple means of communication, including digital platforms; using adequate, simple, and nondiscriminatory language; and taking differences in health literacy into account [27].

Digital Data to Model COVID-19 Spread, Evolution, and Perception

The monitoring of online or social media activities for public health purposes has been investigated since the first days of the digital epidemiology field in the early 2010s, with the objective of capturing health-related trends and modeling disease outbreaks. The most famous examples of internet health surveillance were developed to predict the incidence of influenza, such as Google Flu Trends, or to obtain insights from social media platforms such as Twitter about influenza A/H1N1 [28], measles [29], the Zika virus [30], and the Ebola virus [31,32]. Despite their high potential in public health, these studies have been criticized for their lack of theorization and appropriate standard methodology [33], which can prevent comparison of their results and raises questions regarding the safety of relying on their findings to design targeted public health measures in the real world. To move beyond the simplistic thought that Twitter is a “giant playground” for data scientists, ongoing initiatives to use Twitter data to study the COVID-19 outbreak should build upon previous mistakes, with a more rigorous approach as well as an open source/open data mentality.

Online activities are very good indicators of trends in society. Exploiting data such as Google Trends [34], Wikipedia searches, or the Baidu Index and Weibo Index in China has already proved useful. In China, internet searches and social media data related to COVID-19 have been shown to be highly correlated with daily disease incidence and exhibit an online peak 10 to 14 days before the peak of daily incidences in the real world [35]. Although there are few reports to date, such data sources could also prove useful for early identification of less common symptoms related to COVID-19 using a similar approach to that used to detect weak pharmacovigilance signals from social media and identify side effects of drugs [36]. The main advantage of these data sources is that they can be accessed early in the epidemic and at low cost.

However, predicting disease incidence from online activities or social media data has never been used in a real time prospective fashion for public health guidance on a large scale; rather, such predictions are performed retrospectively by training algorithms to replicate gold-standard data, such as those obtained from the CDC [37], as closely as possible. COVID-19 may be the first outbreak where such a prospective analysis occurs, combining data from standard sources of disease surveillance with social media data or activity trackers [38] to improve overall accuracy.

Social media data analysis is also relevant during the COVID-19 outbreak to monitor how people react to the pandemic evolution over time (sentiment, anxiety, level of stress) as well as common beliefs, opinions [39,40], fears, or hopes regarding treatment or vaccines. Perhaps most importantly, identifying and combating the spread of “fake news” about COVID-19 is highly necessary in our modern digital society [41,42]. This could help to rationalize the debates and actions around COVID-19 and should be further extended [43].

Discussion

Challenges Related to Digital Health Implementation

The COVID-19 crisis is a typical example of the impossibility of establishing a single global technological solution to a given problem. To increase acceptability of digital technologies, the different cultural, moral, and religious backgrounds of users should be considered. With priority given to collective public health benefit and maintaining local social order during the current emergency [44], digital measures can be intrusive and can erode individual freedoms. In some countries, a strong digital divide persists today, and vulnerable populations may be overlooked during the implementation of digital approaches [45-47]. Digital solutions may be less frequently understood and used by people with low health literacy levels or by specific subgroups such as minorities, older individuals, or people who live in rural or low-income areas [48]. For all these reasons, digital approaches can be received in very different ways when they are applied in high-income, medium-income, or low-income areas or when they are deployed in individualistic versus collectivistic countries [49].

In Singapore, an app called TraceTogether has been used to track patients infected with COVID-19. Based on Bluetooth signaling, Singaporean health authorities can track individuals and inform them if they have been in contact with a patient infected with COVID-19 [50]. This patient monitoring system may be beneficial in disease surveillance and outbreak management; however, in European countries, it would not be possible to use it due to individual data protection laws. France conceived the CoronApp, which is based on similar standards [51]. Individuals need to register on the app and provide information on their health status and whether they have symptoms. A geolocation system is used to trace this information and is updated every hour. King's College London conceived the COVID Symptom Tracker app, which aims to help patients monitor their own symptoms [52]. The app gained such popularity that it is now also used in the United States. In addition to its individual benefits, the app can be used for research purposes to study the epidemiology of COVID-19.

Urged by the European Commission, an unprecedented consortium of eight giant telecommunications companies recently agreed to share aggregated and anonymized geolocation data from their clients to track their movements and activities. This raises numerous questions around data privacy [53] and ethics in Western and African countries [54], despite the availability of privacy-preserving proximity tracing technology such as the Decentralized Privacy-Preserving Proximity Tracing (DP-3T) protocol [55]. In Asia, the response to the COVID-19 crisis in South Korea has received much attention and has been widely cited as an exemplary model. With the increased flexibility in South Korea’s digital privacy laws since the MERS crisis in 2015, numerous aggressive digital solutions have been used, such as a national smartphone app, GPS tracking of infected people, and detailed emergency alerts when a positive case has been identified in a person’s geographical area. However, it was proved that this approach was only efficient to reduce the number of new daily cases due to a vast parallel testing campaign to identify COVID-19 hotspots combined with strong isolation and quarantine measures. In Europe, where testing is not as widespread as in South Asia, the use of such monitoring using GPS coordinates is therefore ethically questionable.

Regarding telehealth, the challenge of implementing feasible systems lies with governments in most cases [56], as billing systems must be adapted [15]. Telemedicine consultations must also ensure patient security in terms of data protection. Hence, close collaborations between different actors, such as health care professionals, health technology companies, and health politicians, must be ensured [15]. Once such a system is running, both patients and health care professionals can benefit from better allocation of resources and adequate care, all with reduced risk of exposure. Telemedicine and digital technologies such as apps can also play larger roles in the present crisis not only to fight COVID-19 but also to address the frequent health issues associated with isolation or quarantine, such as psychological needs [57], mental health conditions [58], and physical inactivity

[59]. Lastly, some developing countries face major obstacles to the effective delivery of digital health solutions in rural and remote locations, such as incomplete or insufficient basic digital infrastructures (eg, computers, internet networks, and electricity), lack of sustainable funding to develop, operate, and maintain digital platforms, and high telecommunication costs [4,60,61].

Recommendations to Improve Future Digital Health Systems

In France, where the COVID-19 epidemic is growing at a rapid pace, the health technology ecosystem has rapidly and generously contributed to management of the crisis. Numerous telemedicine and telemonitoring platforms have been locally deployed in hospitals or general practices. However, a lack of coordination between operators has led to issues such as an absence of consensus about the symptoms to monitor and the alert systems to establish. We could argue that some telemonitoring is better than no monitoring; however, given the rapidly evolving body of knowledge around COVID-19, more transparent communication and greater comparability should be pursued to rapidly and homogeneously update the different solutions. Otherwise, some populations will lose the opportunity to gather this data and later use it for research or evaluation.

In previous emergency situations such as the Ebola virus outbreak, Hurricane Harvey, and Hurricane Irma, as well as currently during the COVID-19 pandemic, telehealth has been explored to support patient-physician communication and, hence, improved health outcomes. It must be ensured that outside of emergency situations, telehealth will be appropriately implemented into national health care systems in the long term. This would facilitate use of telehealth systems during outbreaks [62]. Germany is an example of a country where early implementation of telemedicine and mobile apps into the health care sector since 2018 have shown short-term benefits during the COVID-19 pandemic [63]. The German health ministry even claims that they now aim to accelerate the digitization of health care in Germany and help to translate new laws into practical solutions.

Digital initiatives to fight COVID-19 should be relevant to both hospital and public health systems, adapted to the population, rapidly deployable, and capable of evolving with the growth of the body of knowledge related to COVID-19. Regarding common standards for data collection in digital COVID-19 solutions, future initiatives should consider the example of ISARIC (the International Severe Acute Respiratory and emerging Infection Consortium [64]), who successfully organized clinical research on COVID-19 by providing standardized but customizable clinical research forms in several languages, ensuring a harmonized core set of international and comparable data, or the DP-3T initiative, an open implementation of a decentralized privacy-preserving proximity tracing solution [65].

Conclusion

Time is key to fight COVID-19, and digital health solutions provide the opportunity to buy time and human resources. As the COVID-19 pandemic is the first true global health crisis in the digital era, we have observed and will observe a plethora of digital solutions. This pandemic has at least demonstrated the usefulness and reactivity of digital health solutions and constitutes an opportunity to insert these solutions into our health care systems in the long term. This creates an urgent need for policy makers, researchers, and health professionals to collectively and efficiently implement digital solutions into practice without further fragmenting the existing landscapes of care. We now call for more concerted measures to have an optimal impact on the epidemic and to address the most strategic needs to ease the life of people who are at the forefront of the COVID-19 crisis [66].

Acknowledgments

GF and LH are supported by the Luxembourg National Research Fund (FNR) (Predi-COVID, Grant Number 14716273).

Conflicts of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. World Health Organization. 2020 Mar 07. Responding to community spread of COVID-19: interim guidance, 7 March 2020 URL: https://apps.who.int/iris/handle/10665/331421 [accessed 2020-06-08]
3. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 - Studies Needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/NEJMp2002125] [Medline: 32074416]
4. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
5. Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 Dashboard URL: https://coronavirus.jhu.edu/map.html [accessed 2020-04-21]

6. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
7. Power B. The Washington Post. 2020 Mar 27. The coronavirus is expanding the surveillance state. How will this play out? URL: https://www.washingtonpost.com/politics/2020/03/27/coronavirus-is-expanding-surveillance-state-how-will-this-play-out/ [accessed 2020-04-21]
8. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
9. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the Crisis Into an Opportunity: Digital Health Strategies Deployed During the COVID-19 Outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
10. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
11. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
12. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
13. Wang C, Sarsenbayeva Z, Chen X, Dingler T, Goncalves J, Kostakos V. Accurate Measurement of Handwash Quality Using Sensor Armbands: Instrument Validation Study. JMIR Mhealth Uhealth 2020 Mar 26;8(3):e17001 [FREE Full text] [doi: 10.2196/17001] [Medline: 32213469]
14. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
15. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
16. HealthManagement. COVID-19: What Can Healthcare Learn? URL: https://healthmanagement.org/c/healthmanagement/issuearticle/covid-19-what-can-healthcare-learn [accessed 2020-04-22]
17. maladiecoronavirus.fr. URL: https://maladiecoronavirus.fr/ [accessed 2020-06-08]
18. Beat COVID-19 Now. URL: https://beatcovid19now.org/ [accessed 2020-06-08]
19. Martin A, Nateqi J, Gruarin S, Munsch N, Abdarahmane I, Knapp B. An artificial intelligence based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot Internet. bioaRxiv 2020 Apr 06:preprint. [doi: 10.1101/2020.03.25.008805]
20. Wetsman N. The Verge. 2020 Mar 04. Effective communication is critical during emergencies like the COVID-19 outbreak URL: https://www.theverge.com/2020/3/4/21164563/coronavirus-risk-communication-cdc-trump-trust [accessed 2020-04-21]
21. World Health Organization. 2020 Mar 27. Mental health and psychological resilience during the COVID-19 pandemic URL: http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/mental-health-and-psychological-resilience-during-the-covid-19-pandemic [accessed 2020-04-21]
22. World Health Organization. WHO Health Alert brings COVID-19 facts to billions via WhatsApp URL: https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp [accessed 2020-04-21]
23. ICTworks. 2017 Jan 25. How We Fought Ebola with Information - ICTworks Internet URL: https://www.ictworks.org/how-we-fought-ebola-with-information/ [accessed 2020-04-21]
24. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental Health, Risk Factors, and Social Media Use During the COVID-19 Epidemic and Cordon Sanitaire Among the Community and Health Professionals in Wuhan, China: Cross-Sectional Survey. JMIR Ment Health 2020 May 12;7(5):e19009 [FREE Full text] [doi: 10.2196/19009] [Medline: 32365044]
25. Ryan G, De Silva M, Terver JS, Ochi OP, Eaton J. Information systems for global mental health. Lancet Psychiat 2015 May;2(5):372-373. [doi: 10.1016/S2215-0366(15)00097-8] [Medline: 26360265]
26. Basu M. GovInsider. 2020 Mar 03. Exclusive: How Singapore sends daily Whatsapp updates on coronavirus URL: https://govinsider.asia/innovation/singapore-coronavirus-whatsapp-covid19-open-government-products-govtech/ [accessed 2020-04-21]
27. ReliefWeb. 2020 Mar 15. COVID-19: How to include marginalized and vulnerable people in risk communication and community engagement - World Internet URL: https://reliefweb.int/report/world/covid-19-how-include-marginalized-and-vulnerable-people-risk-communication-and [accessed 2020-04-21]
28. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]

29. Du J, Tang L, Xiang Y, Zhi D, Xu J, Song H, et al. Public Perception Analysis of Tweets During the 2015 Measles Outbreak: Comparative Study Using Convolutional Neural Network Models. J Med Internet Res 2018 Jul 09;20(7):e236 [FREE Full text] [doi: 10.2196/jmir.9413] [Medline: 29986843]
30. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts. JMIR Public Health Surveill 2017 Apr 20;3(2):e22 [FREE Full text] [doi: 10.2196/publichealth.6925] [Medline: 28428164]
31. Liang H, Fung IC, Tse ZTH, Yin J, Chan C, Pechta LE, et al. How did Ebola information spread on twitter: broadcasting or viral spreading? BMC Public Health 2019 Apr 25;19(1):438 [FREE Full text] [doi: 10.1186/s12889-019-6747-8] [Medline: 31023299]
32. Morin C, Bost I, Mercier A, Dozon J, Atlani-Duault L. Information Circulation in times of Ebola: Twitter and the Sexual Transmission of Ebola by Survivors. PLoS Curr 2018 Aug 28;10:E1 [FREE Full text] [doi: 10.1371/currents.outbreaks.4e35a9446b89c1b46f8308099840d48f] [Medline: 30254789]
33. Tang L, Bie B, Park S, Zhi D. Social media and outbreaks of emerging infectious diseases: A systematic review of literature. Am J Infect Control 2018 Sep;46(9):962-972 [FREE Full text] [doi: 10.1016/j.ajic.2018.02.010] [Medline: 29628293]
34. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
35. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):pii=2000199 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
36. Tricco AC, Zarin W, Lillie E, Jeblee S, Warren R, Khan PA, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. BMC Med Inform Decis Mak 2018 Jun 14;18(1):38 [FREE Full text] [doi: 10.1186/s12911-018-0621-y] [Medline: 29898743]
37. Alessa A, Faezipour M. Flu Outbreak Prediction Using Twitter Posts Classification and Linear Regression With Historical Centers for Disease Control and Prevention Reports: Prediction Framework Study. JMIR Public Health Surveill 2019 Jun 25;5(2):e12383 [FREE Full text] [doi: 10.2196/12383] [Medline: 31237567]
38. Viboud C, Santillana M. Fitbit-informed influenza forecasts. Lancet Digital Health 2020 Feb;2(2):e54-e55 [FREE Full text] [doi: 10.1016/s2589-7500(19)30241-9]
39. Luo X, Zimet G, Shah S. A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008 - 2017). Hum Vaccin Immunother 2019;15(7-8):1496-1504. [doi: 10.1080/21645515.2019.1627821] [Medline: 31194609]
40. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
41. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
42. Pennycook G, McPhetres J, Zhang Y, Rand D. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention. PsyArXiv 2020 Mar 17:preprint [FREE Full text] [doi: 10.31234/osf.io/uhbk9]
43. Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
44. Pan X. Application of personal-oriented digital technology in preventing transmission of COVID-19, China. Ir J Med Sci 2020 Mar 27:1-2 [FREE Full text] [doi: 10.1007/s11845-020-02215-5] [Medline: 32219674]
45. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
46. Sim I. Mobile Devices and Health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
47. Gamble A, Pham Q, Goyal S, Cafazzo JA. The Challenges of COVID-19 for People Living With Diabetes: Considerations for Digital Health. JMIR Diabetes 2020 May 15;5(2):e19581 [FREE Full text] [doi: 10.2196/19581] [Medline: 32392473]
48. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the Internet as a resource for health information: Results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
49. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
50. Payments News & Mobile Payments Trends. 2020 Mar 25. App Lets Singapore Track Virus Patients' Movements URL: https://www.pymnts.com/coronavirus/2020/app-lets-singapore-track-virus-patients-movements/ [accessed 2020-04-22]

51. Aron M. L'Obs. 2020 Mar 26. « CoronApp », une application française pour géolocaliser les porteurs du coronavirus URL: https://www.nouvelobs.com/coronavirus-de-wuhan/20200326.OBS26600/coronapp-une-application-francaise-pour-geolocaliser-les-porteurs-du-virus-covid-19.html [accessed 2020-04-22]
52. Jee C. MIT Technology Review. 2020 Mar 25. A new app might help researchers monitor the spread of coronavirus URL: https://www.technologyreview.com/2020/03/25/950310/a-new-app-might-help-researchers-monitor-the-spread-of-coronavirus/ [accessed 2020-04-22]
53. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
54. Ekong I, Chukwu E, Chukwu M. COVID-19 Mobile Positioning Data Contact Tracing and Patient Privacy Regulations: Exploratory Search of Global Response Strategies and the Use of Digital Tools in Nigeria. JMIR Mhealth Uhealth 2020 Apr 27;8(4):e19139 [FREE Full text] [doi: 10.2196/19139] [Medline: 32310817]
55. GitHub. DP-3T Documents URL: https://github.com/DP-3T/documents [accessed 2020-05-15]
56. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Apr 02:e82 [FREE Full text] [doi: 10.1056/nejmp2005835]
57. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiat 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
58. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The Role of Telehealth in Reducing the Mental Health Burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
59. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
60. Muinga N, Magare S, Monda J, English M, Fraser H, Powell J, et al. Digital health Systems in Kenyan Public Hospitals: a mixed-methods survey. BMC Med Inform Decis Mak 2020 Jan 06;20(1):2 [FREE Full text] [doi: 10.1186/s12911-019-1005-7] [Medline: 31906932]
61. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the Use of Mobile Health in Improving Health Outcomes in Developing Countries: Systematic Review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
62. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20:1357633X20916567 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
63. Olesch A. Healthcare IT News. 2020 Mar 26. Germany benefits from digital health infrastructure during COVID-19 pandemic URL: https://www.healthcareitnews.com/news/europe/germany-benefits-digital-health-infrastructure-during-covid-19-pandemic [accessed 2020-04-21]
64. International Severe Acute Respiratory and emerging Infection Consortium. COVID-19 New: Clinical data report URL: https://isaric.tghn.org/ [accessed 2020-04-21]
65. GitHub. DP^3T URL: https://github.com/DP-3T [accessed 2020-04-21]
66. Call for papers: COVID-19 research rapidly peer-reviewed and published in JMIR journals. JMIR Publications. URL: https://www.jmir.org/announcement/view/202 [accessed 2020-06-16]

Abbreviations

CDC: US Centers for Disease Control and Prevention
COVID-19: coronavirus disease
DP-3T: Decentralized Privacy-Preserving Proximity Tracing
e-prescription: electronic prescription
ISARIC: International Severe Acute Respiratory and emerging Infection Consortium
MERS: Middle East respiratory syndrome
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 11.04.20; peer-reviewed by SM Ayyoubzadeh, C Auschra; comments to author 20.04.20; revised version received 22.04.20; accepted 04.06.20; published 16.06.20.

Please cite as:
Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L
Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers
J Med Internet Res 2020;22(6):e19284
URL: http://www.jmir.org/2020/6/e19284/
doi:10.2196/19284
PMID:32501804

Review

The Role of Virtual Reality in Improving Health Outcomes for Community-Dwelling Older Adults: Systematic Review

Gordana Dermody[1*], BSN, MSN, DPhil; Lisa Whitehead[1*], BSc, MA, DPhil; Graham Wilson[2*], BA, DPhil; Courtney Glass[1*], BSc, MSc

[1]School of Nursing and Midwifery, Edith Cowan University, Joondalup, Australia

[2]University of Strathclyde, Glasgow, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Gordana Dermody, BSN, MSN, DPhil
School of Nursing and Midwifery
Edith Cowan University
270 Joondalup Drive
Joondalup, 6027
Australia
Phone: 61 8 6304 3470
Fax: 61 8 6304 2323
Email: g.dermody@ecu.edu.au

Abstract

Background: Virtual reality (VR) delivered through immersive headsets creates an opportunity to deliver interventions to improve physical, mental, and psychosocial health outcomes. VR app studies with older adults have primarily focused on rehabilitation and physical function including gait, balance, fall prevention, pain management, and cognition. Several systematic reviews have previously been conducted, but much of the extant literature is focused on rehabilitation or other institutional settings, and little is known about the effectiveness of VR apps using immersive headsets to target health outcomes among community-dwelling older adults.

Objective: The objective of this review was to evaluate the effectiveness of VR apps delivered using commercially available immersive headsets to improve physical, mental, or psychosocial health outcomes in community-dwelling older adults.

Methods: Peer-reviewed publications that included community-dwelling older adults aged ≥60 years residing in residential aged care settings and nursing homes were included. This systematic review was conducted in accordance with the Joanna Briggs Institute (JBI) methodology for systematic reviews of effectiveness evidence. The title of this review was registered with JBI, and the systematic review protocol was registered with the International Prospective Register of Systematic Reviews.

Results: In total, 7 studies that specifically included community-dwelling older adults were included in this review. VR apps using a head-mounted display led to improvements in a number of health outcomes, including pain management, posture, cognitive functioning specifically related to Alzheimer disease, and a decreased risk of falls. A total of 6 studies reported a statistically significant difference post VR intervention, and 1 study reported an improvement in cognitive function to reduce navigational errors. Only one study reported on the usability and acceptability of the interventions delivered through VR. While one study used a distraction mechanism for pain management, none of the studies used gaming technology to promote enjoyment.

Conclusions: Interventions to improve health outcomes through VR have demonstrated potential; however, the ability to synthesize findings by primary outcome for the older adult population is not possible. A number of factors, especially related to frailty, usability, and acceptability, also need to be explored before more substantial recommendations on the effectiveness of VR interventions for older adults can be made.

Trial Registration: PROSPERO CRD42019143504; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=143504

(J Med Internet Res 2020;22(6):e17331) doi:10.2196/17331

KEYWORDS

virtual reality; aged; aged, 80 and over; outcome assessment, health care; independent living; systematic review

Introduction

Background

Understanding how to best support the health and well-being of older adults is an important societal question. With increasing age, natural age-related physical, cognitive, and social changes are often compounded by physical and cognitive comorbidities [1-3]. A primary desire of older people as they age is that they are able to remain living as independently as possible [1-3]. However, the majority of older people live with at least one chronic disease with subsequent impact on physical, mental, or psychosocial health [1-3]. This can impact the ability to live independently or maintain social connectivity [1-3]. Prominent musculoskeletal changes include muscle loss and weakness, which impact gait and balance, increasing the risk of falls [4]. Falls among older adults is a global health issue affecting nearly 35% of people aged 65 years and older every year [5]. As age increases, so does the prevalence of falls [6]. Falls are costly to the health care systems and usually impact the quality of life (QoL) and independence of the older person [7]. Psychological and behavioral responses are common in people who have sustained a fall [8]. In addition, normal neurocognitive changes that occur in healthy aging can impact processing speed and motor responses, potentially increasing the risk of falls [9].

Changes in neurocognitive function are prevalent among older adults. Worldwide, 50 million people are diagnosed with dementia, and Alzheimer disease (AD) is the most common form of dementia, contributing nearly 70% to the dementia diagnoses [10]. The brain changes that accompany AD result in memory loss that can disrupt a person’s daily life [11], for example, difficulty in solving problems and completing everyday tasks at home, misplacing things, and getting lost outside the home, which can have a significant impact on the ability to navigate daily life. Frequently, people who experience memory problems withdraw from social activities because the changes they have experienced impact social interaction [12]. Furthermore, the QoL of older adults with dementia may be adversely affected by chronic pain, which is associated with depression [13]. As people with dementia such as AD may have vision and perception problems, they are at a greater risk of sustaining a fall and three times more likely to suffer from injuries such as hip fracture [14]. Interventions to facilitate healthy aging, adaptation, and strengthening in older adults with normal age-related changes that are intertwined with chronic disease are important in working to support older adults to remain in their own homes for as long as possible and to promote well-being. Exploring interventions that use virtual reality (VR) may offer unique opportunities to address areas of health need. VR is evolving at a rapid rate and presents an opportunity to enhance and support older adults’ physical and cognitive issues to promote engagement in physical activity, travel, or interactions with others [15]. The term *virtual reality* was first coined in the late 1980s. In the last 30 years, a vast range of VR devices have been developed, and the technology has been used across various domains, including education, athlete training, architecture, and notably in health-related research including physical, mental, and psychosocial health [15].

The term *virtual reality* has been used in health research to refer to various forms of onscreen digital representation of real objects or environments, such as showing a silhouette of an individual on a TV screen that mimics their physical rehabilitation movements [16]. However, VR is now predominantly used to refer to what was previously qualified as *immersive virtual reality*, which is understood to be viewing a stereoscopic virtual environment through a motion-tracked head-mounted display (HMD) [16]. A stereoscopic display encompasses the field of view and is able to track the motion of the head [16]. VR using an HMD not only provides the user with a particular artificial environment to view but also allows the user to engage with the virtual environment as they can look around by moving the head and by moving around the virtual environment by moving the body [16]. Definitions of what constitutes *immersion* both within and outside VR research vary, but a concept that arises from motion-tracked VR is *presence*, also known as an illusory experience of being in another place [16]. Deeper conceptualizations of presence include illusions of *place* or being in another location and *plausibility*, where events viewed by a person in VR are convincing [17]. Within the virtual world, the events seen within it are felt as if the user is really there or the events are actually taking place [18]. By allowing the user to interact with the VR environment naturally (eg, walking up to it and using a hand to pick up objects), the experience appears real for the user [18]. These attributes make a number of VR apps such as exposure therapy treatment for anxiety highly effective [18].

The degree of *presence* (or place illusion) that a user experiences is largely dependent on the characteristics of the VR technology being used [17,19]. Earlier VR research utilized HMDs with poor immersive qualities and limited technical capabilities, such as a lack of head-tracking, a small field of view, or the use of transparent displays (something closer to augmented reality) [19]. Greater presence is felt when the VR device has a stereoscopic display, has a wider field of view, and has greater amounts of tracked user movement [19]. Real-world health settings are likely to use these commercial headsets, as they are relatively affordable, easy to set up, and they support VR apps development through software such as Unity or Unreal Engine [20,21]. Although nonimmersive systems are interesting and beneficial, many community-dwelling older adults with natural age-related changes may not have the mental and physical capacity that may be needed in a nonimmersive environment.

Immersive headsets, or those that incorporate motion-tracked stereoscopic HMDs, and potentially motion controllers are known to induce a stronger sense of presence and potentially a sense of realism, embodiment, memory, and spatial understanding than nonimmersive devices [17,19,22,23]. Older adults have natural age-related changes that impact cognitive function; the literature discusses that increasing the immersive properties can be beneficial to increasing the sense of presence and place illusion, which is critical to achieve high intrinsic motivation to participate in the VR experience and achieve the health benefits desired [17,24].

Some community-dwelling older adults with cognitive and physical changes may find it challenging to concentrate on images on a wall-mounted flat screen or a handheld tablet while using an exercise bike. Furthermore, spatial cognition and ability are common changes that occur as a result of aging [25]. Given these issues, the participant may experience increased levels of stress during the intervention, and the sense of presence may be diminished as the participant tries to cognitively and physically navigate the intervention [17,24,25]. The task environment (flat screen, computer screen, and tablet) may have several display elements including passive crystal displays and light-emitting diodes [25]. Lighting levels for reading passive liquid crystal displays (nonbacklit) may be adequate for office or research environments; however, they are not adequate when used with older adults who may have visual acuity changes and who will be using the VR HMD in the home environment [25].

Using nonimmersive displays is further compounded by the angle from which the participant may need to view the display and potentially the glare of the screens that are used, which can alter the ability to visualize [25]. Due to these factors, immersive headsets are considered a distinct category of device with unique potential benefits for community-dwelling older adults. Accordingly, VR experiences in health apps for community-dwelling older adults using HMDs with stereoscopic displays that encompass the field of view and track the motion of at least the head, but also additional parts of the body, were examined in this systematic review.

Therefore, a number of papers have evaluated or systematically reviewed literature based on *virtual reality* in the health of older adults; however, as discussed earlier, the definition of *virtual reality* varies widely. The majority of these papers focused on nonimmersive digital games technology such as the Nintendo Wii, Microsoft Kinect, or custom activities where player movement is shown on a flat TV or monitor screen. Common topics of inquiry into this form of *virtual reality* used with older adults include one or more aspects of physical activity such as balance, rehabilitation, gait function, and falls [26-36]. Serious mental illness [37] and cognitive intervention and training [27,38-40] have also been the focus of reviews; however, very few papers have examined immersive VR, the area of interest in this paper. Duque [41] reported using *3D virtual-reality goggles* for visual-vestibular balance training exercises, although it is not clear if these constitute an immersive display or headset. However, recent research has begun to use immersive HMDs in the assessment of physical [42,43] or cognitive function [44] in older adults as well as providing enjoyable travel experiences [45].

In addition to the effectiveness of VR and digital games on chosen outcomes, it is important to understand the usability and acceptability of technology. When a technology such as VR is used, researchers need to be aware that older adults may not be familiar with the technology or they may have unique aging-related physical or cognitive changes that could make participation difficult. Silva [46] conducted a systematic review to identify the study characteristics of VR research with older adults and how VR impacted older adult end users. Silva [46] concluded that design characteristics could be modified to aid older adult users' experience in using the VR apps. A further systematic review [33] reported evidence on the effectiveness and feasibility of using digital gaming systems within older adults to enable physical activity is weak, with a high risk of bias. Although the authors use the term *virtual reality*, none of the studies used immersive VR HMDs, only traditional flat screen gaming technologies or augmented reality glasses (a transparent display with digital overlay). The acceptance of HMD VR, possible cybersickness, and attitudes toward technology have also been investigated [47]. There was no evidence to suggest that negative attitudes toward VR apps or cybersickness would be barriers to using VR apps with older adults.

Objectives

Although research using VR and immersive headsets is increasing, more knowledge is needed about the types of VR apps using immersive headsets, the health outcomes targeted, outcomes, and transferability to real-world settings for community-dwelling older adults. The aim of this systematic review was to evaluate the effectiveness of VR apps delivered using commercially available immersive headsets in improving physical, mental, and psychosocial health outcomes in community-dwelling older adults.

Methods

Inclusion Criteria

Types of Participants

This review included specifically community-dwelling older adults aged 60 years or older. Initial searches of the existing literature demonstrated that the classification of what constituted *older adult* varied. Accordingly, the United Nations definition of older populations as 60 years and over was used because this classification provided a more realistic representation of older adults across different countries with varying life expectancies [48]. Older adults residing in residential aged care settings and nursing homes were included, and hospitalized older adults were excluded from the review. Studies that included participants aged under 60 years in the control group were also included.

Interventions

Studies that evaluated the use of VR apps delivered via immersive headsets including a combination of VR delivery, employing immersive headsets with incorporation of projector screens to create other forms of immersive VR experiences, were included in this review. Accordingly, commercial headsets including Rift or Vive and older models that encompass the field of view and have motion tracking were included. Studies that did not use VR headsets were excluded.

Comparators

This review considered studies that compared the intervention with conventional therapy and studies that did not use comparisons such as case studies.

Outcomes

This review considered studies that included the following outcomes: effectiveness of the VR apps intervention via immersive headset using physical, mental, or psychosocial outcome measures. These outcomes were measured using kinematic and kinetic computer apps and/or statistical software as well as Likert-based questionnaires.

Types of Studies

This review considered both experimental and quasi-experimental study designs, including, but not limited to, randomized controlled trials (RCTs), nonrandomized controlled trials, before and after studies, and interrupted time-series studies. In addition, analytical observational studies including prospective and retrospective cohort studies, case-control studies, and analytical cross-sectional studies were considered for inclusion. This review also considered descriptive observational study designs including case series, individual case reports, and descriptive cross-sectional studies for inclusion. Human studies published in English between 1997 and 2019 were included. This time frame was chosen to capture the most recent advances of VR technology in health practice.

Review Registration and Approach

The inclusion criteria, methods, and analysis were registered in the Joanna Briggs Institute (JBI) Systematic Review Register, and the protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO, CRD42019143504) a priori to the commencement of data extraction. The systematic review was conducted in accordance with the JBI methodology for systematic reviews of effectiveness evidence. This review followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) recommendations for systematic review reporting [49].

Search Strategy

To identify published studies, a 3-step search strategy was employed between April and June 2019. An initial limited search of the Medical Literature Analysis and Retrieval System Online (MEDLINE, using EbscoHost) and Google Scholar was undertaken to identify articles on the topic, using search terms “virtual reality,” “older adult*,” “community setting,” “health,” and “wellbeing.” The text words contained in the titles and abstracts of relevant articles and the index terms used to describe the articles were used to develop a full search strategy for MEDLINE, Cyberpsychology and Behavior, Cumulative Index of Nursing and Allied Health Literature (CINAHL), Exerpta Medica DataBASE Guide (EMBASE), Web of Science, JBI, the Cochrane Library, IEEExplore, Lancet Psychiatry, Association for Computing Machinery (ACM) Digital Library, Science Direct, and PsycINFO. The search strategy, including all identified keywords and index terms, was adapted for each database to account for appropriate Medical Subject Headings (MeSH) terms. Multimedia Appendix 1 shows the search strategies used for each database, with concepts combined with Boolean operators *AND* and *OR*. Hand searches were conducted when necessary. The reference list of all relevant systematic reviews was screened for additional studies. Multimedia Appendix 2 illustrates the PRISMA flow diagram.

Study Selection

Following the search, all identified citations were collated and uploaded into Endnote X9, and duplicates were removed [50]. Titles and abstracts were screened by 2 independent reviewers for assessment against the inclusion criteria for the review (GD and CG). The full text of selected citations was assessed in detail against the inclusion criteria by at least two independent reviewers (GD, LW, GW, and CG). Reasons for exclusion of full-text studies were as follows: participants in the intervention group being younger than 60 years, VR apps not utilizing HMDs, studies using qualitative designs, and not addressing the effectiveness of VR apps. Disagreements arising between the reviewers at any stage of the study selection process were resolved through team discussions or with a third reviewer.

Assessment of Methodological Quality and Certainty of the Findings

Eligible studies were critically appraised and methodological quality was assessed by 2 independent reviewers (GD, LW, GW, and CG), using the standardized critical appraisal instruments from the JBI [51]. This allowed the reviewers to gain greater insight into the methodological strengths and limitations of the selected studies. Blinding treatment groups was not always conceivable given the nature of the intervention; hence, it was not considered a criterion for inclusion. Any incongruities in appraisal that arose between reviewers were discussed and resolved by all authors. The author of one of the studies considered for inclusion in this review was contacted to request clarification of the randomization techniques used.

The grading of recommendations, assessment, development, and evaluation (GRADE) approach for assessing the certainty of evidence for an effect, summarized in a narrative form, was used to assess the overall quality of the findings [52]. The GRADE assessment evaluates the limitations of the studies, indirectness, imprecision, inconsistency, and publication bias [52]. The overall quality of the evidence was categorized as high, moderate, low, or very low. At least two reviewers (GD, LW, GW, and CG) independently completed the GRADE assessments for each article.

Data Extraction

Data were extracted from studies by 2 independent reviewers (GD, LW, GW, and CG) using an adapted version of the JBI standardized data extraction tool. The data extracted included specific details about the populations; study methods; interventions; outcomes of significance to the review objective, including the effectiveness of the VR intervention on physical, mental, and psychosocial outcomes; types of VR apps used; and specific health outcomes targeted in community-dwelling older adults. Disagreements between the reviewers were resolved through team discussions or with a third reviewer.

Data Synthesis

Findings from the selected studies were narratively synthesized to demonstrate the effectiveness of the VR app intervention via immersive headsets on physical, mental, or psychosocial outcomes.

Results

Summary of the Search Results

The initial search (Multimedia Appendix 1) from 12 databases resulted in 416 references. Additional records were identified (n=74) through manual searching, including searching for systematic review reference lists. After duplicates were removed, titles of a total of 464 studies were assessed.

After the title search was completed, the abstracts of 216 articles were assessed for inclusion by 2 authors (GD and CG). A total of 17 full-text papers were comprehensively assessed for inclusion. Of which, 10 full-text manuscripts were excluded for reasons previously stated, and 7 manuscripts were included in this systematic review, with 2 papers reporting on the same sample. To compare data, Table 1 and Table 2 present the characteristics of the studies and the key findings extracted.

Table 1. Details of study design, virtual reality therapy, and health outcomes.

Study	Study design	Sample size	Sample characteristics	Virtual reality application	Virtual reality exposure	Outcomes assessed	Outcomes measured
Benham [53]	Mixed methods exploratory study	12	8 female, 4 male, majority of the sample were white, majority had exposure to technology weekly, 8 participants had musculoskeletal pain, and 4 had neurological pain	The immersive VR[a] system utilized the HTC[b] Vive HMD[c] with 2 hand controllers. A variety of games could be played, although those using movements or standing were excluded for safety reasons. Popular games included interactions with animals, music, and travel. Used range of commercially available VR apps	Participants completed 12 VR sessions over a 6-week period. These ranged from 15 to 45 min long.	Pain management, QoL[d], depression, overall physical health, overall psychological health, and social life	Patient-reported outcomes, measurement information system, Item Bank v. 1.0–emotional distress–depression, World Health Organization Quality of Life Scale Brief Version (assessing overall health, physical, psychological, and social health), numeric pain rating scale
Gago [54]	Quasi-experimental	39	20 participants with AD[e] (11 fallers and 9 nonfallers) and 19 control participants. Groups were equally matched for demographic and anthropometric data.	Immersive Oculus Rift VR headset showed a view as if the person was standing at the top of a staircase. The perspective in the headset was shifted forward down the stairs to mimic the displacement during a fall. Used the commercially released “Tuscan Villa” demo from Unity, which is no longer available	One session with a total of five 10-second trials per subject, with each trial involving 1 shift in perspective down the stairs (approximate distance of 1.17 m)	Posture and falls in AD	Kinematic and time-frequency distribution sensor data to assess posture and falls

Study	Study design	Sample size	Sample characteristics	Virtual reality application	Virtual reality exposure	Outcomes assessed	Outcomes measured
Levy [55]	RCT[f]	16	10 women and 6 men, randomly assigned to a treatment (6 women and 3 men) and control (4 women and 3 men) group. Their fear of falling developed after falling themselves.	Immersive V8 VR HMD was used with wireless mouse for exposure therapy, followed by the use of serious games. Participants were asked to walk during the VR exposure in different settings on different difficulty levels (city, castle, island, and underground). A 3-dimensional electromagnetic sensor was fixed to the helmet to measure head and body motion during walking periods. Used bespoke research software.	12 weekly sessions lasting 40 min (VR exposure=15 min)	Falls, depression, anxiety, and social life impact from functional impairment	Fear of falling measure, Beck depression inventory, Spielberger state-trait anxiety inventory, and Sheehan disability scale
Optale [56]	RCT	36	24 female and 12 male; from local rest-care home	Compare face-to-face music therapy (control) with VR therapy that includes the music from control condition. Custom virtual environments including outdoor areas with paths, childhood homes, or modern city. Navigated using joystick and viewed through motion-tracked V6 headset. User navigates through the environment and views video clips at certain points. Music plays during the experience. Used bespoke research software.	The experimental group completed 6 months of VR memory training. 3 VR sessions every fortnight within the first 3 months. During the next 3-month booster training phase, there was 1 weekly VR session. VR sessions lasted approximately 15 min.	General cognitive ability and memory, depression, and daily living	General cognitive abilities were assessed using the mini-mental state examination and mental status in neurology tests; the digital span test assessed short-term verbal memory abilities; verbal story recall test assessed long-term verbal memory; phonemic verbal fluency test, dual task performance test, and cognitive estimation test were used to assess executive function; clock drawing test assessed visuospatial processing; activities of daily living functions and mobility and the instrumental activities of daily living assessed daily living activities; the geriatric depression scale assessed depression.

Study	Study design	Sample size	Sample characteristics	Virtual reality application	Virtual reality exposure	Outcomes assessed	Outcomes measured
Parijat [57,58]	RCT	24	12 male and 12 female; 12 control and 12 experimental; closely age matched (±4 years); closely matched for physical characteristics.	Participants viewed a custom city-street VR environment that moved as if the person was walking, through an immersive motion-tracked Sony headset. The visual scene showed that the person was slipping, to induce slip recovery behavior. Used bespoke research software.	One session of VR exposure of 45 to 55 min, with pre- and postslip training and assessments in separate sessions.	Falls	Kinematic and kinetic data were filtered before angular kinematics and muscle activations were assessed at 5-min time intervals.
White [59]	Case study	1	74-year-old male living at home with his wife. Diagnosed with mild cognitive impairment with probable development of AD. Scored 24/30 on Montreal Cognitive Assessment.	Bespoke 3-story virtual building that is navigated via a wheelchair in a 2-dimensional physical space and viewed through oculus Rift DK2. Virtual elevators move the user between floors. Task involves moving to the correct window of 12 in the building. Used bespoke research software.	3×45-min training sessions per week for 7 weeks	Cognitive ability in AD	Cognitive ability and spatial trajectories were assessed

[a] VR: virtual reality.

[b] HTC: high-tech computer.

[c] HMD: head-mounted display.

[d] QoL: quality of life.

[e] AD: Alzheimer disease.

[f] RCT: randomized controlled trial.

Table 2. Summary of findings table.

Outcomes assessed; study	Key findings	Grading of recommendations, assessment, development, and evaluation[a]
Pain management; Benham [53]	Between presession 1 and postsession 12 VR[b] sessions, there were significant improvements in pain scores with a large effect size (−1.54, 95% CI −2.50 to −0.58; *P*=.002; Effect size >0.8=large effect).	Moderate
Quality of life; Benham [53]	WHOQOL-BREF[c] did not find any significant differences over 6 weeks of VR therapy on overall health (−0.06, 95% CI −0.91 to 0.78; *P*=.66), no significant differences on overall physical health (0.41, −0.45 to 1.26; *P*=.08), no significant differences on social life (0.08, 95% CI −0.77 to 0.93; *P*=.87), and no significant differences on overall psychological health (0.33, 95% CI −0.52 to 1.18; *P*=.15).	Moderate
Posture; Gago [54]	The AD[d] faller group had a higher power regarding use of mechanical properties of oscillation for postural adjustments compared with the control group, alluding to worse postural stability in this group (−4 to 0s: *P*=.02; 0 to 4s: *P*=.01; and 4 to 8s: *P*=.008). AD participants had a time lag in cognitive strategies for postural correction compared with healthy subjects (−4 to 0s: *P*=.002; 0 to 4s: *P*=.01).	Moderate
Falls; Gago [54]	The AD fallers groups had a delayed reaction time for changes in power compared with the control group, with a change in power seen only in the last interval (0 to 4s vs 4 to 8s; LB[e]*P*=.008; HB[f]*P*=.01).	Moderate
Falls; Levy [55]	There were statistically significant differences between the 2 groups for fear of falling scores. Fear of falling scores over the 12 weeks reduced by 2.78 (SD 4.82) in the VR group and increased by 4.14 (SD 4.30) in the control group (*P*=.007).	Moderate
Falls; Parijat [57,58]	VR training led to significantly better balance on slippery surfaces with VR therapy reducing slip distance (slip distance 1: $F_{1,18}$=10.34, *P*=.01; slip distance 2: $F_{1,18}$=5.27, *P*=.03), reducing peak slide heel velocity ($F_{1,18}$=4.54, *P*=.05), and reducing peak trunk extension post slip ($F_{1,18}$=12.46, *P*=.01). Slip distance 1 and 2 are the anterior-posterior distance traveled (in cm, based on the location of the heel) from the start of the slip to when heel acceleration peaks (slip distance 1) and then from this point until the heel velocity peaks (distance 2) VR with the treadmill supported realistic walking gait after 15 to 20 min (step duration: $F_{6,76}$=10.56, *P*=.002; step width: $F_{6,76}$=9.56, *P*=.02). There were no significant effects on ankle, hip, or knee kinematics.	Moderate
Memory and cognitive function; Optale [56]	Combining music therapy with exploration of spatial and personally relevant environments in VR led to improved memory ($F_{2,58}$=17.40; *P*<.001) and general cognitive functions (MMSE[g] scores $F_{2,58}$=23.01, *P*<.001; mental status in neurology score $F_{2,58}$=30.16, *P*<.001); executive function (cognitive estimation test group difference: $F_{1,29}$=11.12, *P*=.002; dual task performance test group and time interaction: $F_{2,58}$=10.92, *P*<.001; phonemic verbal fluency test group and time interaction: $F_{2,58}$=14.6, *P*<.001); verbal memory (digital span test group and time interaction $F_{2,58}$=17.4, *P*<.001; verbal story recall test group and time interaction $F_{2,58}$=36.66, *P*<.001), but not in spatial abilities ($F_{2,58}$=3.14, *P*=.05).	Low
Memory and cognitive function; White [59]	Navigation errors reduced during VR training, but there may be no strong positive effect on overall cognitive ability.	Low
Disability; Levy [55]	There were no significant differences in social life (*P*=.18) or family life (*P*=.12) impact from functional impairment between the VR exposure therapy group and the waiting-list group.	Moderate
Daily living; Optale [56]	There were no significant differences of VR therapy on daily living tasks (Activities of Daily living $_{2,58}$=1.5, *P*=.23; Instrumental Activities of Daily Living $F_{2,58}$=1.05, *P*=.36).	Moderate
Anxiety; Levy [55]	There were statistically significant differences between the 2 groups' mean state anxiety scores. The mean state anxiety score reduced by 8.86 (SD 14.46) in the VR group and increased by 9.80 (4.66) in the control group (*P*=.005). There were no significant differences in trait anxiety scores between the groups (*P*=.24).	Moderate
Depression; Benham [53]	There were no significant differences in PROMIS[h] scores (0.29, 95% CI −1.14 to 0.56; *P*=.33).	Moderate
Depression; Levy [55]	There were no significant differences in Beck Depression Inventory scores between the VR and control groups (*P*=.47).	Moderate

Outcomes assessed; study	Key findings	Grading of recommendations, assessment, development, and evaluation[a]
Depression; Optale [56]	The participants receiving VR therapy had a reduced depression value on the Geriatric Depression Scale after the initial VR session ($F_{1,29}$=5.61; P=.02), but not after the booster VR session ($F_{1,29}$=1.35; P=.25).	Moderate

[a] GRADE assessment reported per outcome, not per study.

[b] VR: virtual reality.

[c] WHOQOL-BREF: World Health Organization Quality of Life Scale Brief Version.

[d] AD: Alzheimer disease.

[e] LB: low-frequency band within kinematic time-frequency analysis.

[f] HB: high-frequency band within kinematic time-frequency analysis.

[g] MMSE: mini-mental state examination.

[h] PROMIS: patient-reported outcomes measurement information system.

Populations Included in the Review

The populations recruited were highly variable. Although all were 60 years and older, the populations recruited into the studies in this review were older adults experiencing musculoskeletal or neurological pain, with a fear of falling, living in residential aged care, diagnosed with AD, and diagnosed with mild cognitive impairment [53-59]. Therefore, the results could not be synthesized across defined populations, for example, older adults at risk of falls.

Virtual Reality Application

VR apps were used to view a stereoscopic virtual environment through a motion-tracked HMD. Each of the studies in this review used an immersive virtual environment intervention [53-59]. This included differently manufactured devices with varying frames per second, field of view, degrees of freedom, latency, and tracking, which influence the immersive sense of presence [19]. The frequency and duration of VR exposure varied considerably between studies. VR exposure ranged from five 10-second displays of stereoscopic 3D images to 55 min of serious gaming, with 15 min being the most common exposure time [53-59]. VR therapy was administered as a singular session in 3 studies [54,57,58]. Other studies completed VR therapy more regularly, with Levy [55] administering it weekly over a 12-week period, Benham [53] completing it twice weekly for 6 weeks, White [59] completing it 3 times per week over a 7-week period, and Optale [56] completing 3 sessions per fortnight. Across the 7 papers, VR was used for 2 general categories of health care application: *treatment* for a particular condition or symptom and *assessment* of physical ability (although there was also a component of training) [53-59]. A total of 2 papers used VR apps that were commercially available at the time: Benham [53] allowed participants to use apps from VivePort, an web-based software portal (www.viveport.com), whereas Gago [54] made use of the *Tuscan Villa* demonstration environment that was provided in the Unity game engine software (which is no longer available) [20]. All other papers utilized bespoke software designed and created for the purposes of the research.

The papers in the review shared few commonalities, as a number of different health outcomes were addressed and a range of VR headsets and apps were utilized. Despite this, VR was often found to be effective or beneficial in addressing health in older adults, as it has with pain and physical activity [53-59]. This suggests that it can be a flexible and effective means of intervention delivery.

Health Outcomes Targeted in the Interventions

Physical Health Outcomes

Physical health outcomes are important for community-dwelling older adults to remain living in the community as independently as possible. However, due to the complex multimorbid chronic disease, many community-dwelling older people are at risk of declining physical health as they age [60,61]. For example, older adults living in the community are at greater risk for falls due to increasing functional decline and subsequent muscle loss and weakness. Several studies focused on falls, including participants who were either at risk of falls or with a history of falls [54,55,57,58]. Some overlap in health outcomes existed where 2 studies focused on health outcomes that transect physical, mental, and social aspects [53,55]. For example, the paper by Levy [55] had a primary objective focused on assessing VR therapy in the treatment of fear of falling. In total, 2 studies evaluated the use of VR training apps to improve postural and muscular adjustment with the goal of enabling the older person to compensate for induced perturbation (balance-recovery) or slips with the goal of fall prevention [54,57]. In perturbation-balance training, a balance-recovery approach is used to improve the recovery reaction to a person losing his or her balance. Gago [54] developed VR training that introduced visual perturbation with the goal of studying the postural adjustment mechanisms in participants with AD with and without a history of falls and a control group. A key finding of this study was that older adults with AD who had a history of falling needed greater compensatory postural adjustment compared with participants with AD who did not have a history of falling and the control group (−4 to 0s: P=.02; 0 to 4s: P=.01; and 4 to 8s: P=.008) [54]. This may be due to visual and perceptual problems that are often common in people with AD. However, the participants with AD who had a history of falls were able to make adjustments to visual perturbation [54].

Parijat et al [57,58] conducted 2 separate studies using the same sample and dataset. In the study by Parijat et al [58], the design and effectiveness of VR training to improve recovery reactions of healthy older adults, with the aim of reducing fall frequency, was investigated. Parijat et al [58] found that VR training of motor skills to recover from a slip-induced fall led to significantly better balance minutes (step duration: $F_{6,76}$=10.56, P=.002; step width: $F_{6,76}$= 9.56, P=.02) and recovery reactions (slip distance 1: $F_{1,18}$=10.34, P=.01; slip distance 2: $F_{1,18}$=5.27, P=.03; peak slide heel velocity: F1,18=4.54, P=.05) on actual slippery surfaces. However, because participants were able to adapt to virtual slips fairly quickly, visually inducing physical responses may not be effective long term. In the paper, slip distances 1 and 2 are the anterior-posterior distance traveled (in cm, based on the location of the heel) from the start of the slip to when heel acceleration peaks (slip distance 1) and then from this point until the heel velocity peaks (slip distance 2) [58]. In the study by Parijat et al [57], the objective was to improve the recovery reaction and improve angular and muscular responses when participants were exposed to a slippery surface. The authors also quantified the kinematics of angular and muscular changes [57]. Kinematics is a way of describing the geometric mechanics of motion and velocity without considering the force that caused the motion. The key findings of this study were that VR with the treadmill supported both realistic walking gait and significant improvements in slip recovery kinematics [57]. The authors concluded that slip training was more effective using VR training compared with conventional movable platform training [57]. The kinematic angular differences varied only in the trunk measurements between the 2 groups. Peak trunk extensions decreased more in the VR training group compared with the control group ($F_{1,18}$=12.46; P=.01) [57]. In the second slip trial, the VR training group was able to quickly reverse the forward trunk rotation, which is a key ability for regaining balance [57].

Mental and Psychosocial Health Outcomes

A total of 3 studies implemented VR to improve cognition, memory, and/or psychologic aspects in older adults [55,56,59]. In the study by Levy [55], older adult participants who exhibited fear of falling were included. The objective of this study was to use a VR app to treat the pathology of the phobic reactions to walking experienced by the participants [55]. Notably, all participants had moderate to severe social limitations as a result of their traumatic fall history, in addition to having a variety of comorbidities that were not related to walking difficulties [55]. The findings of this study showed a significant decrease in the fear of falling measure in the intervention group (−2.78; SD 4.82) compared with the control group after VR exposure (4.14; SD 4.30; P=.007) [55]. Although there was no difference in social life or depression, anxiety in the intervention group was significantly lower (−8.86; SD 14.46) than that in the control group (9.80; SD 4.66; P=.005) [55]. Optale [56] and White [59] studied the use of VR apps to improve cognition and memory function in older adults with memory deficits. Optale [56] used VR immersion and interaction as a treatment intervention for older adults with memory deficits. Several outcomes were measured, including general cognitive abilities, verbal memory, executive functions, and visuospatial processing. The study findings show that VR therapy led to significant improvements in cognitive functioning and verbal memory (pretraining to posttraining MMSE score: $F_{1,29}$=6.85; P=.01 and posttraining to postbooster evaluation: $F_{1,29}$=4.46; P=.04), particularly long-term memory improved in the intervention group [56]. A significant change was observed in the experimental group for all 3 executive functions measured [56]. However, visuospatial abilities or daily living were not affected [56]. A descriptive case study by White [59] used VR app intervention to strengthen the cognitive reserve in one participant with early AD to improve and maintain cognition, and in particular, spatial cognition. The participant's navigational ability improved in response to the training given [59]. Although navigation errors were present, a reduction was evident overall [59]. However, it is unclear whether or not there is a positive effect on overall cognitive ability [59]. The findings of this case study suggest that people with early stages of AD can learn to navigate paths in a suitable immersive VR system [59].

In the remaining study, Benham [53] conducted a mixed methods study at a senior day-center to identify the efficacy of VR interventions on pain, depression, and QoL in community-dwelling older adults who self-reported acute or chronic pain that was bothersome at least two days per week. The VR therapy appears to have improved pain management in community-dwelling older adults [53]. VR intervention significantly reduced the report of pain (−1.54, 95% CI−2.50 to −0.58; P=.002) [53]. Participants also reported that the VR intervention was able to distract them from their pain [53].

Discussion

Principal Findings

The incorporation of VR in health care is promising in its ability to support older adults in managing age-related changes, for example, musculoskeletal changes and those related to chronic disease that impact physical, cognitive, and psychosocial health and well-being. The studies included in this review are examples of how technologic advances are changing the face of health care and demonstrate that certain VR interventions can be successfully used with older adults.

Fall prevention appears to be an area where immersive VR using HMD can have a significant impact. Falls contribute to injury, pain, disability, and premature death [5]. The loss of balance and slow reactions to steady state is a primary mechanism for falls [62]. A total of 2 studies evaluated the use of VR training apps to improve postural and muscular adjustment with the goal of enabling the older person to compensate for induced perturbation (balance-recovery) or slips so that a fall may be prevented [54,57]. Although there is an emphasis on the physical aspect of fall prevention, including balance, gait, and recovery reaction, sustaining a fall can have psychological and behavioral impacts on the older person. Fear of falling is common among older adults, and in some cases, this fear is disproportionate, with physical reactions to extreme anxiety including palpitations, sweating, and avoidance of walking altogether [63,64]. A total of 3 studies aimed to improve cognition, memory, and/or psychological aspects in older adults [55,56,59]. The results indicate a positive change and improvement, although several questions remain around the longevity of the impact [55,56,59].

The findings of the studies reviewed suggest that the expansion of VR for health-related interventions, while in a safe and controlled environment, may have several potential therapeutic benefits for older adults to facilitate independence and QoL [65,66]. Although the studies included in this review have mostly focused on evaluating the effectiveness of VR intervention in conventional therapy [53-59], VR intervention apps could also support therapists and clinicians to assess a person's physical, cognitive, or psychosocial status to identify potential problems [67,68]. For example, gait patterns, postural adjustment, musculoskeletal pain, and compensatory mechanisms could be assessed using VR to identify mobility issues of concern and determine a person's fall risk [69]. Tailored VR interventions could be developed to target fall prevention and to promote increased physical activity, which has a protective effect on age-related musculoskeletal changes [54,57,58].

Furthermore, using VR to assess, develop, and implement cognitive interventions to improve cognition, memory, and psychological aspects could make VR a competitor to traditional cognitive rehabilitation [70]. For example, stroke is common among older adults, and many require cognitive rehabilitation along with physical rehabilitation. Clinicians could use VR as an individualized intervention to determine cognitive and decision-making problems and develop individualized VR-based interventions [70,71]. Chronic pain is common in older adults, limiting independence, physical movement, and social activity [53,72]. Although more knowledge is needed about VR interventions to reduce or manage pain in older adults, VR has been used extensively in other age groups, and there is scientific evidence that VR can reduce the pain that is experienced [73,74]. VR exposure therapy interventions have been applied to reduce extreme fears such as phobias [73-75]. After a fall or similar accident, older adults often suffer from extreme fear of falling, which could be very limiting to staying physically active [76]. The study by Levy [55] was able to demonstrate significant reductions in the fear of falling. One of the unique benefits of VR use in health care with older adults is the potential to meet more than one need, as VR apps may span the physical, mental, and psychosocial needs of a person [77,78].

Although some studies under review reported indication of positive outcomes regarding VR use in older adults [53,55,59], it is important to highlight several practical aspects that are challenges to implementing VR interventions with older adults, who are often in a vulnerable state, and to translate the use of VR interventions as a standard of care in the aged care sector. The majority of participants were healthy enough to fully participate in the intervention studies [57-59]. Investigations into using VR interventions with older adults who are frail or unwell are needed. For example, a better understanding of the severity of cognitive impairment and the applicability of VR has not been explored. People with increased levels of frailty who are at greater risk for functional decline and falls may benefit from VR interventions; however, it is unclear whether frail older adults can engage with the VR interventions described in this review.

Strengths and Limitations

Visual and auditory changes could make it challenging for some older adults to engage with the technology. It is important to consider natural age-related changes in the design of VR apps for older adults [79]. Only Parijat et al [57,58] explicitly assessed visual acuity, via Snellen's chart; Benham [53], Gago [54], Levy [55], and Optale [56] excluded participants if they self-reported uncorrected vision or serious/significant sensory impairment. None of the authors administered hearing tests [53-59]. Only Benham [53] and Parijat et al [57,58] took any measurement of cybersickness [80]. In most of the studies, it was unclear whether the researchers used a participatory approach by inviting older adults to contribute to the design and conduct of the intervention [81]. Although some literature indicates that older adults are open to using VR interventions, not much is known about the acceptance of HMD VR use in older adults [66,82]. For example, only 1 study included a survey question about the overall experience with VR [53]. None of the other studies discussed how well older adults accepted the VR and HMD and whether they were able to engage with and adapt to the physical aspects of the VR device and the VR environment or their experience with *presence*. Unless there is some adaptation to vision and hearing abnormalities and dexterity, older adults may be reluctant to adopt VR as a consistent health care–related intervention. In using VR exposure therapy with older people with anxiety and phobias, there is some debate on the outcome expectancy of treatment outcomes induced by *presence* (or immersion in the VR environment) because this relationship could be influenced by anxiety that may be inadvertently created by *presence* [55]. Subsequently, clinicians and industry providers may be reluctant to adopt VR for exposure therapy.

We noted that several studies in this review used a VR intervention as a means to impact another outcome that was not measured at the time of the study [54,57,58]. For example, although some of the studies implemented VR training with the aim of preventing falls, fall prevention was an indirect outcome variable, and although postural adjustment may have improved during the study, it is unclear how long this benefit lasts for and whether any longer-term impact on falls was experienced. Furthermore, some authors noted that VR interventions that induce slips need to make each slip novel in some way to reduce the likelihood that participants adapt to the stimulus and inadvertently limit the desired postural improvement [57,58].

Interestingly, no study designed or incorporated gaming technology in VR to support engagement and promote enjoyment for participants [26]. This would bring the technology *full circle* with its beginnings as a gaming technology [83]. The participants in the study by Benham [53] were able to play a game if they desired, but many seemed to choose other activities such as traveling or interacting with animals, and gaming/gamification was not a deliberate or controlled part of the intervention design or evaluation. Longitudinal studies should be conducted to determine the long-term effect on the outcome variable under study. All studies in this review included a small sample size, which impacted the generalizability of the findings [53-59]. Powered RCTs are needed to determine the effectiveness of VR therapy compared with standard care practices.

It is important to note that none of the papers included in the review reached a *High* level of GRADE certainty and quality. White [59] was judged to have a very low GRADE judgment, as it constituted a single case study with no blinding or randomization. The remaining papers were rated as *Moderate*, as they employed randomized designs, blinding, patient-reported outcomes, or RCTs [53-58]. Therefore, there is convincing evidence that VR interventions can be effective; however, it is clear that more research needs to be done. A recent review paper highlighted a similar shortcoming in scientific rigor in other fields of VR research [84]. So, although this suggests that it is not a problem particular to the health domain, it also emphasizes that more rigorous scientific methods need to be used to robustly evaluate and validate the technology.

Conclusions

In conclusion, the wider literature suggests that VR interventions have the potential for wide-scale adoption to promote health and wellness among older adults. This review demonstrates that interventional research using VR with older adults in varying states of health is in the early stages of development. In particular, evidence around the potential adoption of VR interventions in applied clinical or therapeutic settings is limited and requires further understanding of logistics, financial costs, and the acceptability and usability of immersive VR in older adults from the perspective of older adults and clinicians. The review also indicated the need for a greater understanding of the design features of immersive VR apps that can promote and improve health for older adults, including those with audio-visual deficits.

Acknowledgments

This study was not funded. GD and CG completed the title and abstract search. Review of full-text papers, data extraction, interpretation of results, and writing of the manuscript was completed by all authors (GD, LW, GW, and CG).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 20 KB - jmir_v22i6e17331_app1.docx]

Multimedia Appendix 2
Preferred reporting items for systematic reviews and meta-analyses flow diagram.
[PNG File , 132 KB - jmir_v22i6e17331_app2.png]

References

1. Pipher M. Another Country. New York, USA: Berkley Publishing Group; 1999.
2. National Seniors Australia. 2012. Where Will I Live as I Age? Senior Australians' Needs and Concerns about Future Housing and Living Arrangements URL: https://nationalseniors.com.au/uploads/121201_NationalSeniorsResearch_GroupHomesAus_WhereWillILive.pdf [accessed 2019-10-14]
3. Pew Research Center. 2016. Most Older Adults Want to Age in Place URL: http://www.pewsocialtrends.org/2016/02/18/smaller-share-of-women-ages-65-and-older-are-living-alone/st_2016-02-18_older-adult_0-02/ [accessed 2019-10-30]
4. Sherrington C, Fairhall N, Wallbank G, Tiedemann A, Michaleff Z, Howard K, et al. Exercise for preventing falls in older people living in the community: an abridged Cochrane systematic review. Br J Sports Med 2019 Dec 2:- epub ahead of print. [doi: 10.1136/bjsports-2019-101512] [Medline: 31792067]
5. World Health Organisation. 2019. Falls Prevention in Older Age URL: https://www.who.int/ageing/projects/falls_prevention_older_age/en/ [accessed 2019-11-01]
6. Gale C, Cooper C, Sayer AA. Prevalence and risk factors for falls in older men and women: the English longitudinal study of ageing. Age Ageing 2016 Nov;45(6):789-794 [FREE Full text] [doi: 10.1093/ageing/afw129] [Medline: 27496938]
7. Bjerk M, Brovold T, Skelton D, Bergland A. A falls prevention programme to improve quality of life, physical function and falls efficacy in older people receiving home help services: study protocol for a randomised controlled trial. BMC Health Serv Res 2017 Aug 14;17(1):559 [FREE Full text] [doi: 10.1186/s12913-017-2516-5] [Medline: 28806904]
8. Suzuki M, Ohyama N, Yamada K, Kanamori M. The relationship between fear of falling, activities of daily living and quality of life among elderly individuals. Nurs Health Sci 2002 Dec;4(4):155-161. [doi: 10.1046/j.1442-2018.2002.00123.x] [Medline: 12406202]
9. Harada C, Love MC, Triebel K. Normal cognitive aging. Clin Geriatr Med 2013 Nov;29(4):737-752 [FREE Full text] [doi: 10.1016/j.cger.2013.07.002] [Medline: 24094294]
10. World Health Organisation. 2019. Dementia URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2019-11-01]

11. Dementia Australia National. 2017. Alzheimer's Disease URL: https://www.dementia.org.au/about-dementia/types-of-dementia/alzheimers-disease [accessed 2019-11-01]
12. Alzheimer's Association: Alzheimer's Disease & Dementia Help. 2019. Alzheimer's Disease Facts and Figures URL: https://www.alz.org/alzheimers-dementia/facts-figures [accessed 2019-11-01]
13. Rostad H, Puts M, Småstuen MC, Grov E, Utne I, Halvorsrud L. Associations between pain and quality of life in severe dementia: a Norwegian cross-sectional study. Dement Geriatr Cogn Dis Extra 2017;7(1):109-121 [FREE Full text] [doi: 10.1159/000468923] [Medline: 28553313]
14. Baker N, Cook M, Arrighi H, Bullock R. Hip fracture risk and subsequent mortality among Alzheimer's disease patients in the United Kingdom, 1988-2007. Age Ageing 2011 Jan;40(1):49-54. [doi: 10.1093/ageing/afq146] [Medline: 21087990]
15. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
16. Skarbez R, Brooks Jr FP, Whitton M. A survey of presence and related concepts. ACM Comput Surv 2018 Jan 12;50(6):1-39 [FREE Full text] [doi: 10.1145/3134301]
17. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3549-3557 [FREE Full text] [doi: 10.1098/rstb.2009.0138] [Medline: 19884149]
18. Chesham R, Malouff J, Schutte N. Meta-analysis of the efficacy of virtual reality exposure therapy for social anxiety. Behav Change 2018 Jun 18;35(3):152-166 [FREE Full text] [doi: 10.1017/bec.2018.15]
19. Cummings J, Bailenson J. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol 2015 May 15;19(2):272-309 [FREE Full text] [doi: 10.1080/15213269.2015.1015740]
20. Unity Real-Time Development Platform. 2019. Unity 2019.2 URL: https://unity.com/releases/2019-2?_ga=2.140078651.924856950.1574747078-1810818056.1574747078 [accessed 201-10-14]
21. Unreal Engine. 2019. Frequently Asked Questions (FAQ) URL: https://www.unrealengine.com/en-US/faq [accessed 2019-10-14]
22. Ruddle R, Volkova E, Bülthoff H. Walking improves your cognitive map in environments that are large-scale and large in extent. ACM Trans Comput-Hum Interact 2011 Jun;18(2):1-20 [FREE Full text] [doi: 10.1145/1970378.1970384]
23. Schöne B, Wessels M, Gruber T. Experiences in virtual reality: a window to autobiographical memory. Curr Psychol 2017 Jul 20;38(3):715-719 [FREE Full text] [doi: 10.1007/s12144-017-9648-y]
24. Bruun-Pedersen J, Serafin S, Kofoed L. Going Outside While Staying Inside: Exercise Motivation with Immersive vs Non–Immersive Recreational Virtual Environment Augmentation for Older Adult Nursing Home Residents. In: Proceedings of the 2016 IEEE International Conference on Healthcare Informatics. 2016 Presented at: ICHI'16; October 4-7, 2016; Chicago, IL, USA. [doi: 10.1109/ichi.2016.31]
25. Fisk A, Czaja SJ, Rogers WA, Charness N, Sharit J. Designing for Older Adults: Principles and Creative Human Factors Approaches. Second Edition. Boca Raton, Florida: CRC Press Taylor and Francis Group; 2009.
26. Molina K, Ricci N, de Moraes SA, Perracini M. Virtual reality using games for improving physical functioning in older adults: a systematic review. J Neuroeng Rehabil 2014 Nov 15;11:156 [FREE Full text] [doi: 10.1186/1743-0003-11-156] [Medline: 25399408]
27. Bleakley C, Charles D, Porter-Armstrong A, McNeill M, McDonough S, McCormack B. Gaming for health: a systematic review of the physical and cognitive effects of interactive computer games in older adults. J Appl Gerontol 2015 Apr;34(3):NP166-NP189. [doi: 10.1177/0733464812470747] [Medline: 24652863]
28. Booth V, Masud T, Connell L, Bath-Hextall F. The effectiveness of virtual reality interventions in improving balance in adults with impaired balance compared with standard or no treatment: a systematic review and meta-analysis. Clin Rehabil 2014 May;28(5):419-431. [doi: 10.1177/0269215513509389] [Medline: 24188913]
29. de Amorim JS, Leite R, Brizola R, Yonamine C. Virtual reality therapy for rehabilitation of balance in the elderly: a systematic review and META-analysis. Adv Rheumatol 2018 Jul 31;58(1):18. [doi: 10.1186/s42358-018-0013-0] [Medline: 30657081]
30. Dockx K, Bekkers E, van den Bergh V, Ginis P, Rochester L, Hausdorff J, et al. Virtual reality for rehabilitation in Parkinson's disease. Cochrane Database Syst Rev 2016 Dec 21;12:CD010760 [FREE Full text] [doi: 10.1002/14651858.CD010760.pub2] [Medline: 28000926]
31. Laver K, Lange B, George S, Deutsch J, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
32. Neri S, Cardoso J, Cruz L, Lima R, de Oliveira RJ, Iversen M, et al. Do virtual reality games improve mobility skills and balance measurements in community-dwelling older adults? Systematic review and meta-analysis. Clin Rehabil 2017 Oct;31(10):1292-1304. [doi: 10.1177/0269215517694677] [Medline: 28933612]
33. Miller K, Adair B, Pearce A, Said C, Ozanne E, Morris M. Effectiveness and feasibility of virtual reality and gaming system use at home by older adults for enabling physical activity to improve health-related domains: a systematic review. Age Ageing 2014 Mar;43(2):188-195. [doi: 10.1093/ageing/aft194] [Medline: 24351549]
34. Ravenek K, Wolfe D, Hitzig S. A scoping review of video gaming in rehabilitation. Disabil Rehabil Assist Technol 2016 Aug;11(6):445-453. [doi: 10.3109/17483107.2015.1029538] [Medline: 25815680]

35. Skjæret N, Nawaz A, Morat T, Schoene D, Helbostad J, Vereijken B. Exercise and rehabilitation delivered through exergames in older adults: an integrative review of technologies, safety and efficacy. Int J Med Inform 2016 Jan;85(1):1-16. [doi: 10.1016/j.ijmedinf.2015.10.008] [Medline: 26559887]
36. van Diest M, Lamoth C, Stegenga J, Verkerke G, Postema K. Exergaming for balance training of elderly: state of the art and future developments. J Neuroeng Rehabil 2013 Sep 25;10:101 [FREE Full text] [doi: 10.1186/1743-0003-10-101] [Medline: 24063521]
37. Välimäki M, Hätönen HM, Lahti M, Kurki M, Hottinen A, Metsäranta K, et al. Virtual reality for treatment compliance for people with serious mental illness. Cochrane Database Syst Rev 2014 Oct 8(10):CD009928. [doi: 10.1002/14651858.CD009928.pub2] [Medline: 25300174]
38. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
39. García-Betances RI, Waldmeyer MT, Fico G, Cabrera-Umpiérrez MF. A succinct overview of virtual reality technology use in Alzheimer's disease. Front Aging Neurosci 2015;7:80 [FREE Full text] [doi: 10.3389/fnagi.2015.00080] [Medline: 26029101]
40. Ge S, Zhu Z, Wu B, McConnell E. Technology-based cognitive training and rehabilitation interventions for individuals with mild cognitive impairment: a systematic review. BMC Geriatr 2018 Sep 15;18(1):213 [FREE Full text] [doi: 10.1186/s12877-018-0893-1] [Medline: 30219036]
41. Duque G, Boersma D, Loza-Diaz G, Hassan S, Suarez H, Geisinger D, et al. Effects of balance training using a virtual-reality system in older fallers. Clin Interv Aging 2013;8:257-263 [FREE Full text] [doi: 10.2147/CIA.S41453] [Medline: 23467506]
42. Arias P, Robles-García V, Sanmartín G, Flores J, Cudeiro J. Virtual reality as a tool for evaluation of repetitive rhythmic movements in the elderly and Parkinson's disease patients. PLoS One 2012;7(1):e30021 [FREE Full text] [doi: 10.1371/journal.pone.0030021] [Medline: 22279559]
43. Eisapour M, Cao S, Domenicucci L, Boger J. Participatory Design of a Virtual Reality Exercise for People with Mild Cognitive Impairment. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI EA'18; April 21-26, 2018; Montreal, QC, Canada. [doi: 10.1145/3170427.3174362]
44. Ouellet E, Boller B, Corriveau-Lecavalier N, Cloutier S, Belleville S. The virtual shop: a new immersive virtual reality environment and scenario for the assessment of everyday memory. J Neurosci Methods 2018 Jun 1;303:126-135 [FREE Full text] [doi: 10.1016/j.jneumeth.2018.03.010] [Medline: 29581009]
45. Srifar D. 360 virtual reality travel media for elderly. Int J Math Appl 2018 Jun 30;10(1/2/3):45-52. [doi: 10.5121/ijma.2018.10304]
46. Silva RS, Mol A, Ishitani L. Virtual reality for older users: a systematic literature review. Int J Virtual Real 2019 Jan 1;19(1):11-25 [FREE Full text] [doi: 10.20870/ijvr.2019.19.1.2908]
47. Huygelier H, Schraepen B, van Ee R, Vanden Abeele V, Gillebert CR. Acceptance of immersive head-mounted virtual reality in older adults. Sci Rep 2019 Mar 14;9(1):4519 [FREE Full text] [doi: 10.1038/s41598-019-41200-6] [Medline: 30872760]
48. United Nations. 2017. World Population Ageing URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf [accessed 2019-10-14]
49. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
50. EndNote | Clarivate Analytics. 2019. URL: https://endnote.com/ [accessed 2019-02-01]
51. Joanna Briggs Institute. 2017. Critical Appraisal Tools URL: https://joannabriggs.org/ebp/critical_appraisal_tools [accessed 2019-08-12]
52. Cochrane Handbook for Systematic Reviews of Interventions. 2011. URL: https://handbook-5-1.cochrane.org/ [accessed 2019-08-15]
53. Benham S, Kang M, Grampurohit N. Immersive virtual reality for the management of pain in community-dwelling older adults. OTJR (Thorofare N J) 2019 Apr;39(2):90-96. [doi: 10.1177/1539449218817291] [Medline: 30595096]
54. Gago M, Yelshyna D, Bicho E, Silva H, Rocha L, Rodrigues ML, et al. Compensatory postural adjustments in an Oculus virtual reality environment and the risk of falling in Alzheimer's disease. Dement Geriatr Cogn Dis Extra 2016;6(2):252-267 [FREE Full text] [doi: 10.1159/000447124] [Medline: 27489559]
55. Levy F, Leboucher P, Rautureau G, Komano O, Millet B, Jouvent R. Fear of falling: efficacy of virtual reality associated with serious games in elderly people. Neuropsychiatr Dis Treat 2016;12:877-881 [FREE Full text] [doi: 10.2147/NDT.S97809] [Medline: 27143889]
56. Optale G, Urgesi C, Busato V, Marin S, Piron L, Priftis K, et al. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. Neurorehabil Neural Repair 2010 May;24(4):348-357. [doi: 10.1177/1545968309353328] [Medline: 19934445]
57. Parijat P, Lockhart T, Liu J. EMG and kinematic responses to unexpected slips after slip training in virtual reality. IEEE Trans Biomed Eng 2015 Feb;62(2):593-599 [FREE Full text] [doi: 10.1109/TBME.2014.2361324] [Medline: 25296401]

58. Parijat P, Lockhart T, Liu J. Effects of perturbation-based slip training using a virtual reality environment on slip-induced falls. Ann Biomed Eng 2015 Apr;43(4):958-967 [FREE Full text] [doi: 10.1007/s10439-014-1128-z] [Medline: 25245221]
59. White PJ, Moussavi Z. Neurocognitive treatment for a patient with Alzheimer's disease using a virtual reality navigational environment. J Exp Neurosci 2016;10:129-135 [FREE Full text] [doi: 10.4137/JEN.S40827] [Medline: 27840579]
60. May C, Eton D, Boehmer K, Gallacher K, Hunt K, MacDonald S, et al. Rethinking the patient: using burden of treatment theory to understand the changing dynamics of illness. BMC Health Serv Res 2014 Jun 26;14:281 [FREE Full text] [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
61. Mair FS, May CR. Thinking about the burden of treatment. Br Med J 2014 Nov 10;349:g6680. [doi: 10.1136/bmj.g6680] [Medline: 25385748]
62. Lockhart T, Smith J, Woldstad J. Effects of aging on the biomechanics of slips and falls. Hum Factors 2005;47(4):708-729 [FREE Full text] [doi: 10.1518/001872005775571014] [Medline: 16553061]
63. Harding S, Gardner A. Fear of falling. Aust J of Adv Nurs 2009;27(1):94 [FREE Full text]
64. Stahl S, Albert S. Understanding disproportionate fear of falling in older adults: implications for intervention development. Am J Geriatr Psychiatry 2018 Aug;26(8):860-861 [FREE Full text] [doi: 10.1016/j.jagp.2018.05.003] [Medline: 29859738]
65. Kim O, Pang Y, Kim JH. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. BMC Psychiatry 2019 Jul 12;19(1):219 [FREE Full text] [doi: 10.1186/s12888-019-2180-x] [Medline: 31299921]
66. Syed-Abdul S, Malwade S, Nursetyo A, Sood M, Bhatia M, Barsasella D, et al. Virtual reality among the elderly: a usefulness and acceptance study from Taiwan. BMC Geriatr 2019 Aug 19;19(1):223 [FREE Full text] [doi: 10.1186/s12877-019-1218-8] [Medline: 31426766]
67. Rizzo A, Shilling R. Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. Eur J Psychotraumatol 2017;8(sup5):1414560 [FREE Full text] [doi: 10.1080/20008198.2017.1414560] [Medline: 29372007]
68. Riva G, Wiederhold BK, Mantovani F. Neuroscience of virtual reality: from virtual exposure to embodied medicine. Cyberpsychol Behav Soc Netw 2019;22(1):82-96 [FREE Full text] [doi: 10.1089/cyber.2017.29099.gri]
69. Stokes HE, Thompson JD, Franz JR. The neuromuscular origins of kinematic variability during perturbed walking. Sci Rep 2017 Apr 11;7(1):808 [FREE Full text] [doi: 10.1038/s41598-017-00942-x] [Medline: 28400615]
70. Rogers J, Duckworth J, Middleton S, Steenbergen B, Wilson P. Elements virtual rehabilitation improves motor, cognitive, and functional outcomes in adult stroke: evidence from a randomized controlled pilot study. J Neuroeng Rehabil 2019 May 15;16(1):56 [FREE Full text] [doi: 10.1186/s12984-019-0531-y] [Medline: 31092252]
71. Shin H, Kim K. Virtual reality for cognitive rehabilitation after brain injury: a systematic review. J Phys Ther Sci 2015 Sep;27(9):2999-3002 [FREE Full text] [doi: 10.1589/jpts.27.2999] [Medline: 26504344]
72. Dueñas M, Ojeda B, Salazar A, Mico J, Failde I. A review of chronic pain impact on patients, their social environment and the health care system. J Pain Res 2016;9:457-467 [FREE Full text] [doi: 10.2147/JPR.S105892] [Medline: 27418853]
73. Chi B, Chau B, Yeo E, Ta P. Virtual reality for spinal cord injury-associated neuropathic pain: systematic review. Ann Phys Rehabil Med 2019 Jan;62(1):49-57 [FREE Full text] [doi: 10.1016/j.rehab.2018.09.006] [Medline: 30312665]
74. Mallari B, Spaeth E, Goh H, Boyd B. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019;12:2053-2085 [FREE Full text] [doi: 10.2147/JPR.S200498] [Medline: 31308733]
75. Wechsler T, Kümpers F, Mühlberger A. Inferiority or even superiority of virtual reality exposure therapy in phobias?-a systematic review and quantitative meta-analysis on randomized controlled trials specifically comparing the efficacy of virtual reality exposure to gold standard exposure in agoraphobia, specific phobia, and social phobia. Front Psychol 2019;10:1758 [FREE Full text] [doi: 10.3389/fpsyg.2019.01758] [Medline: 31551840]
76. Jefferis BJ, Iliffe S, Kendrick D, Kerse N, Trost S, Lennon LT, et al. How are falls and fear of falling associated with objectively measured physical activity in a cohort of community-dwelling older men? BMC Geriatr 2014 Oct 27;14:114 [FREE Full text] [doi: 10.1186/1471-2318-14-114] [Medline: 25348492]
77. Hughes S, Warren-Norton K, Spadafora P, Tsotsos L. Supporting optimal aging through the innovative use of virtual reality technology. Multimodal Technol Interact 2017 Sep 28;1(4):23 [FREE Full text] [doi: 10.3390/mti1040023]
78. Dudley D. AARP. 2018. Virtual Reality Used to Combat Isolation and Improve Health URL: https://www.aarp.org/home-family/personal-technology/info-2018/vr-explained.html [accessed 2020-10-21]
79. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. JMIR Serious Games 2018 Oct 17;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]
80. Weech S, Varghese J, Barnett-Cowan M. ScienceDaily. 2018. Virtual Reality Motion Sickness May Be Predicted and Counteracted URL: https://www.sciencedaily.com/releases/2018/09/180927083336.htm [accessed 2019-11-25]
81. Flobak E, Wake J, Vindenes J, Kahlon S, Nordgreen T, Guribye F. Participatory Design of VR Scenarios for Exposure Therapy. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-7, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300799]
82. Roberts A, de Schutter B, Franks K, Radina M. Older adults' experiences with audiovisual virtual reality: perceived usefulness and other factors influencing technology acceptance. Clin Gerontol 2019;42(1):27-33. [doi: 10.1080/07317115.2018.1442380] [Medline: 29505343]

83. Kickstarter. 2019. Oculus Rift: Step Into the Game URL: https://www.kickstarter.com/projects/1523379957/oculus-rift-step-into-the-game [accessed 2019-10-22]
84. Lanier M, Waddell T, Elson M, Tamul D, Ivory J, Przybylski A. Virtual reality check: statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments. Comput Human Behav 2019 Nov;100:70-78 [FREE Full text] [doi: 10.1016/j.chb.2019.06.015]

Abbreviations

AD: Alzheimer disease
GRADE: grading of recommendations, assessment, development, and evaluation
HMD: head-mounted display
JBI: Joanna Briggs Institute
MEDLINE: Medical Literature Analysis and Retrieval System Online
PRISMA: preferred reporting items for systematic reviews and meta-analyses
QoL: quality of life
RCTs: randomized controlled trials
VR: virtual reality

Edited by G Eysenbach; submitted 06.12.19; peer-reviewed by D Lin, F Pallavicini, T Alessa; comments to author 13.01.20; revised version received 05.03.20; accepted 10.04.20; published 01.06.20.

Please cite as:
Dermody G, Whitehead L, Wilson G, Glass C
The Role of Virtual Reality in Improving Health Outcomes for Community-Dwelling Older Adults: Systematic Review
J Med Internet Res 2020;22(6):e17331
URL: https://www.jmir.org/2020/6/e17331
doi:10.2196/17331
PMID:32478662

Review

Therapeutic Alliance in Technology-Based Interventions for the Treatment of Depression: Systematic Review

Eileen Wehmann[1], MPhil; Moritz Köhnen[1], MSc; Martin Härter[1], MD, PhD, Dipl Psych; Sarah Liebherz[1], PhD, Dipl Psych

[1] Department of Medical Psychology, Center for Psychosocial Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
Sarah Liebherz, PhD, Dipl Psych
Department of Medical Psychology
Center for Psychosocial Medicine
University Medical Center Hamburg-Eppendorf
Martinistrasse 52
Hamburg, 20246
Germany
Phone: 49 40 7410 57969
Email: s.liebherz@uke.de

Abstract

Background: There is growing evidence that technology-based interventions (TBIs) are effective for the treatment of depression. As TBIs are gaining acceptance, a question arises whether good therapeutic alliance, considered a key aspect of psychotherapy, can be established without or with minimal face-to-face contact or rather changes if blended concepts are applied. While therapeutic alliance has been studied extensively in the context of face-to-face therapy, only few studies have reviewed evidence on alliance ratings in TBIs.

Objective: The purpose of this study was to examine therapeutic alliance in technology-based psychological interventions for the treatment of depression.

Methods: We searched Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, PsycINFO, PSYNDEX, CINAHL, clinical trial registers, and sources of grey literature for randomized controlled trials on TBIs in the treatment of adults with unipolar depression. All publications were selected according to prespecified criteria. Data were extracted by two independent reviewers.

Results: A total of eight out of 98 studies (9.5%) included in the review on TBIs for depression considered therapeutic alliance as part of their evaluation. The available data covered eight different treatment conditions, including four stand-alone treatments (face-to-face psychotherapy, email, telephone, and internet program) and four combined treatments (face-to-face psychotherapy plus a smartphone app and an internet program combined with face-to-face psychotherapy, treatment as usual, or email/telephone). On average, patients rated the alliance positively across all groups. Importantly, no relevant group differences regarding therapeutic alliance sum scores were found in any of the studies. Five studies investigated the relationship between patients' alliance ratings and treatment outcome, revealing mixed results.

Conclusions: Our results suggest that it is possible to establish a positive therapeutic alliance across a variety of different TBIs for depression, but this is based on a small number of studies. Future research needs to determine on what basis therapeutic alliance is formed in settings that do not allow for additional nonverbal cues, perhaps with adapted instruments to measure therapeutic alliance.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42016050413; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42016050413)

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-028042

(J Med Internet Res 2020;22(6):e17195) doi:10.2196/17195

KEYWORDS

therapeutic alliance; depression; technology-based intervention; treatment

Introduction

Background

Globally, over 300 million people (point prevalence in 2015) are estimated to have depression, accounting for 4.4% of the world population [1]. Depression is the greatest contributor for both disability and suicide [1] and is associated with greatly increased health care costs [2]. Treatment of depression with face-to-face psychotherapy has been shown to be effective over the last decades [3]. There is growing evidence that technology-based interventions (TBIs) are also effective for the treatment of depression, with great clinical improvement and recovery after treatment [4-9].

TBIs differ according to various aspects, including the type of technology (eg, computer and smartphone), therapeutic rationale (eg, cognitive behavior therapy and behavioral activation), the clinical phase of depression management (waiting period, acute treatment, and aftercare), and the amount of human support [10]. The degree of support as well as the delivery mode can range from a stand-alone self-administered treatment with no therapist contact at all to a blended treatment with active involvement of a therapist and regular face-to-face meetings.

Besides lowering the threshold to access treatment (eg, providing geographic and time-related flexibility) [11], TBIs, such as online therapy, provide an opportunity to reach people who refuse to seek traditional services, especially those who wish to remain anonymous [11,12].

Considering these aspects as well as the increasing interest in TBIs, a question arises whether adequate therapeutic alliance, considered a key aspect of psychotherapy, can be established without or with minimal face-to-face contact or rather changes when blended concepts are applied.

Therapeutic alliance can be viewed as the “quality of partnership and mutual collaboration between a therapist and client” [13], albeit a generally valid definition of the construct does not exist. According to Bordin [14], good therapeutic alliance is characterized by an agreement between the therapist and patient on the goals of the therapy, the tasks to be performed, and their emotional bond (eg, if the patient feels respected and appreciated by the therapist). Depending on the type of technology, the communication between the therapist and patient can be synchronous (telephone or video conference) or asynchronous (email or web-based program). Moreover, different technologies provide different cues about the therapist, such as visual (video conference) and auditory (telephone) cues, which may influence the emotional bond. This complexity needs to be incorporated when describing findings concerning therapeutic alliance, as pointed out in a recent review by Berger [15].

There are different instruments to assess therapeutic alliance, and the most frequently used instrument is the Working Alliance Inventory (WAI) [16,17], which includes the three aspects of *goal*, *task*, and *bond*. These instruments can be administered at different time points and are based on the patient’s, the therapist’s, or an observer’s point of view.

Positive alliance ratings are good predictors of treatment success in traditional face-to-face psychotherapy, even when controlling for other possible confounders, such as prior symptom change [13,18,19].

While therapeutic alliance has been studied extensively in the context of face-to-face therapy [13], only few studies have reviewed the evidence on alliance ratings in TBIs [17,20]. In a meta-analysis of 295 studies, Flückiger et al [17] recently showed that alliance ratings in internet-based programs are similar to those in face-to-face settings. They further revealed that the effect size of the relationship between alliance and treatment outcome is comparable in TBIs and face-to-face settings. The authors reported that most of the studies investigating alliance in internet-based interventions relied on the WAI. However, this review focused on internet-based interventions only and did not refer specifically to people diagnosed with depression. There are randomized controlled trials (RCTs) addressing therapeutic alliance in TBIs for people with depression [21,22]. However, there is no published systematic review on this topic.

Aim

The purpose of this study was to examine therapeutic alliance in TBIs for the treatment of depression in a systematic review, considering the following research questions: (1) How many of the studies included in the review consider therapeutic alliance in their evaluation? (2) How is therapeutic alliance assessed? (3) How is the quality of therapeutic alliance rated across different interventions? (4) Is there a relationship between therapeutic alliance and treatment outcome?

Methods

Design

This study is part of a preregistered systematic review (PROSPERO registration number: CRD42016050413) on the comparative effectiveness of *T*echnology-based *I*nterventions in different steps of *De*pression *Ca*re (TIDECA). The corresponding protocol has been published [23] and includes more detailed descriptions. This review is in accordance with the standards of the Cochrane Collaboration [24] and is reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [25].

Search Strategy

We searched the following key databases: Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, PsycINFO, PSYNDEX, and CINAHL. The search was not limited by date, language, or publication status. Supplementary material of the study protocol [23] includes applied search strategies of key databases. We further searched clinical trial registers (ClinicalTrials.gov, International Clinical Trial Registry Platform, and German Clinical Trial Register) and sources of grey literature (Open Grey, Trip Database, ProQuest Dissertations & Theses Abstract and Indexing, and specialized registers of Institute for Scientific Information Web of Science). Additionally, all first authors of the included publications were contacted for supplementary information on further published and unpublished trials and specific study information or the status of on-going studies, which were identified as published study protocols or preregistered trials.

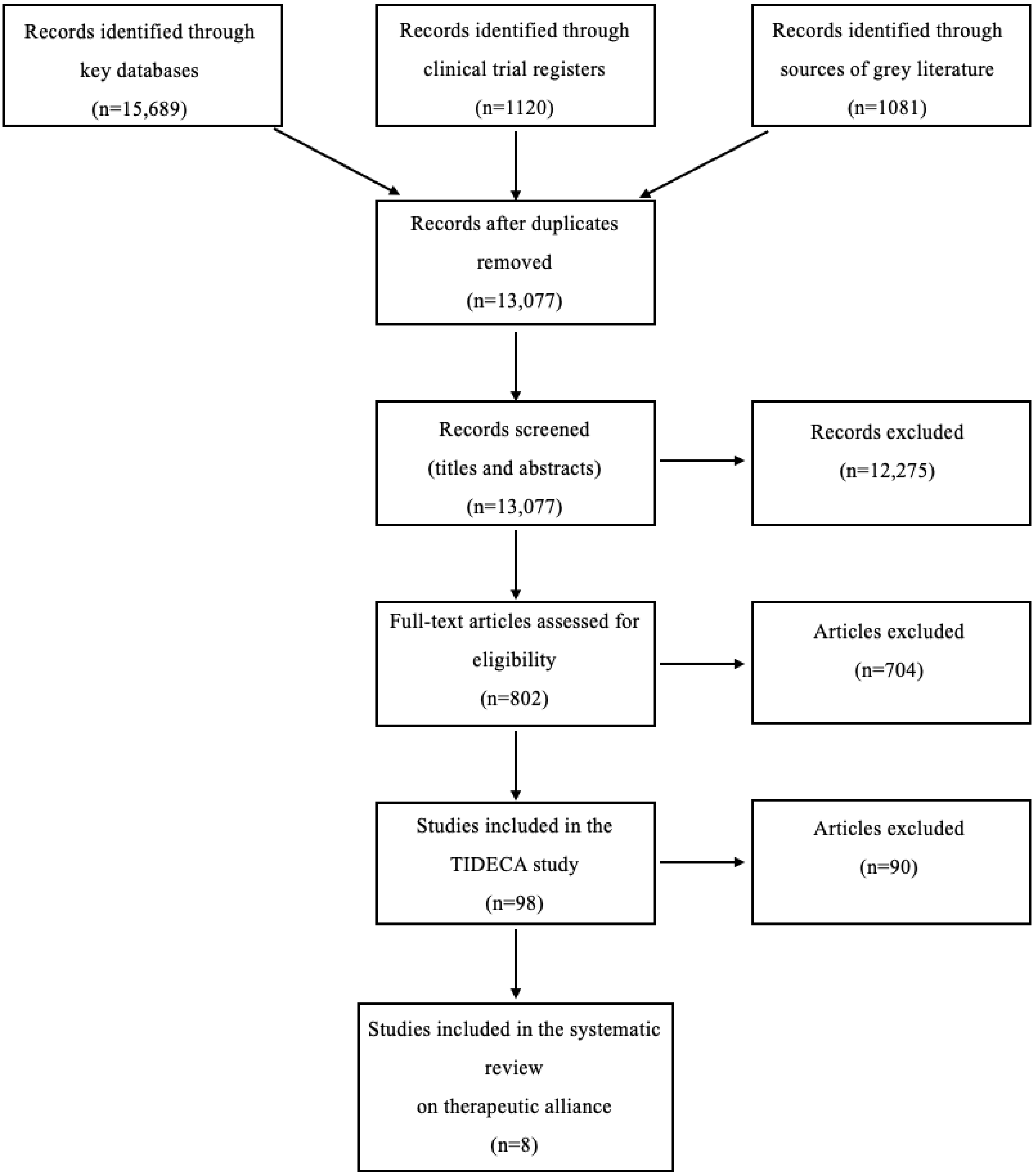
Inclusion and Exclusion Criteria

We applied PICOS categories (Population; Intervention; Comparator; Outcome; Study design) [24,26] to define the inclusion and exclusion criteria for our study. All abstracts were screened according to prespecified criteria. We included studies only if (1) they were cluster RCTs, (2) participants were 18 years or older, (3) participants had a diagnosis of unipolar depression based on a formal classification system, (4) mental or somatic comorbidities were not the main focus of the intervention or study, and (5) the active intervention included a TBI based on an explicit psychotherapeutic theory and aimed to improve depressive symptoms. For more details, refer to the study protocol [23].

Selection Procedure

The selection process is presented in Figure 1. The search yielded at total of 13,077 records after duplicates were removed. Two authors (MK and SL) independently screened the first 100 records for inclusion. Since the interrater reliability for this sample was high (98%), one author (MK) screened the remaining records in the course of the title/abstract screening and the second author (SL) assessed publications labelled as “unclear.” Selected full-text publications (n=802) were subsequently screened for inclusion by two independent reviewers (MK and MD). Discrepancies were resolved by discussion with a third reviewer (SL). A total of 98 publications fulfilled all inclusion criteria for the TIDECA study and were finally screened to determine whether they included data on the *therapeutic working alliance relationship*. Eight publications were included in this review.

Figure 1. PRISMA flow diagram. TIDECA: Technology-based Interventions in different steps of Depression Care.



Data Extraction

We developed a standardized data collection sheet, which was piloted on three studies. The sheet collected the following data: (1) general data (eg, year of publication), (2) methodological information (including risk of bias assessment), (3) demographic and clinical sample characteristics, (4) classification of the clinical phase of depression management (waiting period, acute treatment, and aftercare), (5) delivery mode (eg, blended treatment), (6) treatment characteristics, (7) sample size and study flow, (8) primary and secondary outcome data, and (9) data on therapeutic alliance, including measuring instrument, rating of therapeutic alliance for all included groups (means, standard deviations, and *P* values for available group differences), and data on the relationship between therapeutic alliance and treatment outcomes (*P* values). Data were extracted by two independent reviewers (EW and MK), and inconsistencies were resolved by discussion or by involving a third reviewer (SL).

Data Analysis

Given the heterogeneity of the studies included, qualitative (respectively descriptive) synthesis of the evidence was performed, rather than a meta-analysis.

Descriptive statistics were used to present means, standard deviations, and percentage scores. Standard deviations were derived from standard errors, if not provided otherwise. We decided to focus on alliance ratings after the treatment had started (instead of pretreatment ratings) in order to detect possible group differences based on the different interventions. If several time points after the start of treatment were reported in the studies, we selected the earliest alliance ratings, in line with previous findings showing that therapeutic alliance is established at an early stage of therapy [27]. Given that only a small number of studies provided the therapists' ratings of therapeutic alliance and all studies provided the patients' ratings, the results focused on the patients' ratings.

Quality Appraisal

The risk of bias assessment was evaluated by two independent reviewers (EW and MK) for the included studies according to the Risk of Bias Tool of the Cochrane Handbook for Systematic Reviews [24] (including the domains random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other bias). In line with a previous operationalization [28], we specified the domain "other bias" using the following three additional categories: insufficient treatment adherence, allegiance bias, and attention bias. Interrater reliability was calculated to be 77%. Discrepancies were resolved by discussions between the reviewers (EW and MK) or by involving a third reviewer (SL).

Protocol Changes

Since the applied literature search was very comprehensive, we did not perform an additional systematic forward and backward reference search.

Results

Study Characteristics

Among the selected studies in the TIDECA review, 9.5% (8/98) considered therapeutic alliance as part of their evaluation. As displayed in Table 1, studies examining therapeutic alliance were published within the last decade and were mainly located in Europe, except for one study from the United States [29-36]. The sample size of these studies ranged from 38 to 325 participants. The technological delivery modes in the intervention groups included email, web-based programs, telephone, and a smartphone app. Waitlists, face-to-face psychotherapy, treatment as usual (TAU) combined with a waitlist, and an active control group with TAU were used as controls. Among the studies, four used a TBI as a stand-alone intervention, three implemented a blended treatment (ie, combining TBIs with face-to-face therapy), and one conducted enhanced stand-alone interventions (ie, combining TBIs with TAU without following a specific concept). The therapeutic rationale of applied TBIs was based on behavioral psychotherapy in all studies. The length of the treatment ranged from 8 to 18 weeks. The qualifications of the therapists varied from MSc students of clinical psychology and licensed psychotherapists to PhD-level psychologists.

Notable differences were observed with regard to the degree of therapist guidance. Therapists in predominantly self-help interventions focused on providing feedback, validation, reinforcement, and encouragement to continue with the program. In predominantly therapist-administered interventions, the therapy was delivered by therapists, whereas interventions based on self-help did not involve any therapists.

Table 1. Study characteristics.

Study, authors (publication year)	Study location	N[a]	Study arms	Delivery mode	Therapeutic rational of TBI[b]	Length (weeks)	Qualification of the therapists	Degree of guidance[c]
Andersson et al (2012) [29]	Sweden	88	Email support vs internet program vs waitlist	Stand-alone intervention	CBT[d]	8	MSc students of clinical psychology	Predominantly therapist administered (email) and predominantly self-help (internet)
Berger et al (2018) [30]	Germany	98	Internet program + F2F[e] vs F2F	Blended treatment	CBT	12	Licensed psychotherapists	Predominantly therapist administered
Lindner et al (2014) [31]	Sweden	38	internet program + telephone support vs internet program + email support	Stand-alone intervention	BA[f] + ACT[g]	8	MSc students of clinical psychology	Predominantly self-help
Ly et al (2015) [32]	Sweden	93	Smartphone + F2F vs F2F	Blended treatment	BA	9 (blended) and 10 (control)	MSc students of clinical psychology	Predominantly self-help
Meyer et al (2015) [33]	Germany	163	Internet program + TAU[h] vs TAU + waitlist	Enhanced stand-alone intervention	CBT	13.05	No therapists involved	Self-administered
Steinmann et al (2019) [34]	Germany	59	Telephone + F2F + letters vs telephone + F2F	Stand-alone intervention	CBT	9-13	Licensed psychotherapists	Predominantly therapist administered
Stiles-Shields et al (2014) [35]	USA	325	Telephone vs F2F	Stand-alone intervention	CBT	18	PhD-level psychologists	Predominantly therapist administered
Zwerenz et al (2017) [36]	Germany	229	Internet program + TAU vs active control + TAU	Blended treatment	CBT	12	No therapists involved	Self-administered

[a] N: number of participants randomized.

[b] TBI: technology-based intervention.

[c] Based on the study by Newman et al [37].

[d] CBT: cognitive behavioral therapy.

[e] F2F: regular face-to-face psychotherapy.

[f] BA: behavioral activation.

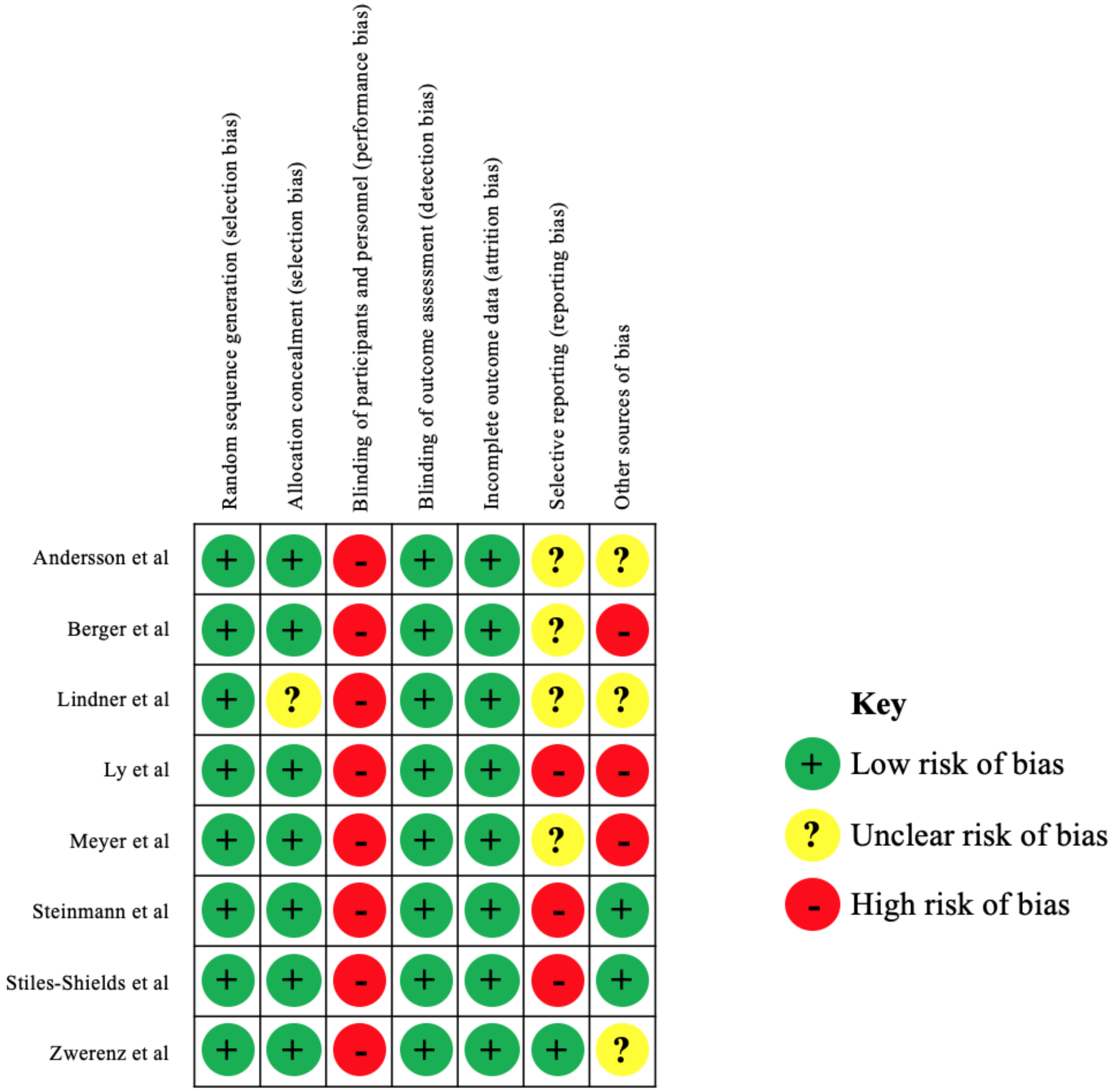
[g] ACT: acceptance and commitment therapy.

[h] TAU: treatment as usual.

Risk of Bias Assessment

Overall, the risk of bias assessment showed a low risk of bias for selection bias, detection bias, and attrition bias (Figure 2). Given the nature of the studies, blinding of participants and personnel was not possible, thus creating a high risk for performance bias. Selective reporting (reporting bias) was unclear or high in all but one study, mainly due to unexplained or unjustified discrepancies between the protocol or the trial registration and the reported measures. Other sources of bias were unclear or high in all but two studies.

Figure 2. Risk of bias assessment. Other sources involved insufficient treatment adherence, allegiance bias or conflict of interest, and attention bias.



	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other sources of bias
Andersson et al	+	+	-	+	+	?	?
Berger et al	+	+	-	+	+	?	-
Lindner et al	+	?	-	+	+	?	?
Ly et al	+	+	-	+	+	-	-
Meyer et al	+	+	-	+	+	?	-
Steinmann et al	+	+	-	+	+	-	+
Stiles-Shields et al	+	+	-	+	+	-	+
Zwerenz et al	+	+	-	+	+	+	?

Key

+ Low risk of bias
? Unclear risk of bias
- High risk of bias

Sample Characteristics

Table 2 provides an overview of the sociodemographic characteristics of the sample. The sample is rather typical for psychotherapeutic studies, with a mean participant age of 43.9 (SD 13.9) years and the majority of participants being female (769/1,093, 70.4%), being employed (554/941, 58.9%), and having a high level of education (672/1034, 65.0%). All studies included only participants who had a clinical diagnosis of depression, except for the study by Meyer et al [33], which had a sample with 84% (137/163) of patients having depression.

Table 2. Sample sociodemographic characteristics.

Characteristic	Value[a] (N=1093[b])
Age (years)	43.9 (13.9)
Sex (females)	769 (70.4%)
Occupation	
Employed (full and part time)	554 (58.9%)
Unemployed (including retired and disabled)	303 (29.3%)
Undergoing training (students and trainees)	41 (9.0%)
Education	
Low educational level (<9 years)	42 (4.1%)
Middle educational level (9-11 years)	297 (28.7%)
High educational level (≥12 years)	672 (65.0%)
Marital status	
Married/with partner	480 (46.4%)
Single/without partner (including divorced and widowed)	329 (46.4%)

[a] Data are presented as mean (SD) or n (%).

[b] Values refer to the available data. Some studies did not report in line with our subcategories (missing data: occupation, four studies; education, one study; marital status, two studies).

Assessment of Therapeutic Alliance

Table 3 summarizes how therapeutic alliance was assessed in the included studies. Five studies [29-32,35] relied on the WAI in its original format of 36 items (7-point Likert scale) [16], the short form version (WAI-S) with 12 items (7-point Likert scale) [38], or the short form-revised version (WAI-SR) with only positively worded items (5-point Likert scale) [39]. All five studies reported the total score, and three studies [29,30,35] further reported the subscores of the WAI (task, goal, and bond).

Two studies relied on the original Helping Alliance Questionnaire [40]; one reported the total score [36], whereas the other reported the subscore for collaboration or bond [34]. One study [33] assessed therapeutic alliance using the 11-item Helping Alliance Questionnaire (HAQ-11) [41], with a 6-point response score. The mean total score of the HAQ-11 was converted to a percentage score that indicated how many participants viewed the alliance as positive.

All studies used the client-rated version of the assessment tools, and three studies [30,35,36] also utilized therapist ratings. While all studies scheduled an early or mid-treatment assessment of the alliance, ranging from 2 to 6 weeks of treatment, some studies included pretreatment and follow-up evaluations as well. Notably, Steinmann et al [34] reported alliance ratings at study intake only, and Zwerenz et al [36] administered the HAQ to measure therapeutic alliance in the TAU arm but not in the TBI arm. Therefore, these data were not considered in the subsequent results.

Table 3. Assessment of therapeutic alliance.

Study, authors (publication year)	Assessment tool	Measures reported	Rater	Time of assessment	Total treatment length
Andersson et al (2012) [29]	WAI[a]	Total score + 3 subscores (task, goal, and bond)	Client	After 3 weeks	8 weeks
Berger et al (2018) [30]	WAI-Short Revised	Total score + 3 subscores (task, goal, and bond)	Client + therapist	After 6/12 weeks	12 weeks
Lindner et al (2014) [31]	WAI-Short Form	Total score	Client	After 2/8/12 weeks	8 weeks
Ly et al (2015) [32]	WAI-Short Form	Total score	Client	Pre-treatment/after 3 weeks	9 weeks (10 weeks control)
Meyer et al (2015) [33]	HAQ-11[b]	Total score	Client	After 3 weeks	13.05 weeks[c]
Steinmann et al (2019) [34]	HAQ[d]	Subscore (collaboration or bonding)	Client	Study intake	9-13 weeks
Stiles-Shields et al (2014) [35]	WAI-Short Revised	Total score + 3 subscores (task, goal, and bond)	Client + therapist	At 4/14 weeks	18 weeks
Zwerenz et al (2017) [36]	HAQ	Total score	Client + therapist	Study intake/end of TAU[e]	12 weeks

[a] WAI: Working Alliance Inventory.

[b] HAQ-11: 11-item Helping Alliance Questionnaire.

[c] A period of 3 months was converted to weeks with a factor of 4.35 weeks per month for comparability.

[d] HAQ: Helping Alliance Questionnaire.

[e] TAU: treatment as usual.

Ratings of Therapeutic Alliance

The patients' ratings of therapeutic alliance (between week 2 and 6) are displayed in Table 4. The available data cover eight different treatment conditions, including four *stand-alone treatments* (face-to-face psychotherapy, email, telephone, and internet program) and four *combined treatments* (face-to-face psychotherapy plus a smartphone app and an internet program combined with face-to-face psychotherapy, treatment as usual, or email/telephone).

On average, patients rated the alliance positively across all groups. Importantly, no relevant group differences regarding therapeutic alliance sum scores were found in any of the studies. Andersson et al [29] found a significant difference (P=.04) between email-based therapy and the use of a self-help internet program when comparing the subscore *goal*, with higher scores in the email condition. However, no relevant group differences regarding the other subscores where found.

Berger et al [30] and Stiles-Shields et al [35] further analyzed therapists' alliance ratings and found no statistically significant group differences between face-to-face psychotherapy and face-to-face psychotherapy combined with an internet program (P>.09) or between face-to-face psychotherapy and telephone treatment (P>.05).

Table 4. Patients' ratings for therapeutic alliance.

Study, authors (publication year) and subscale	Alliance ratings[a], mean (SD)								Group differences[d]
	F2F[b]	F2F + smartphone	Internet program	Internet program + F2F	Internet program + TAU[c]	Email	Internet program + email/phone	Telephone	
Andersson et al (2012) [29]									
Total	N/A[e]	N/A	5.25 (0.82)	N/A	N/A	5.58 (0.82)	N/A	N/A	>.05
Task	N/A	N/A	5.19 (0.84)	N/A	N/A	5.23 (0.83)	N/A	N/A	>.05
Bond	N/A	N/A	5.47 (0.97)	N/A	N/A	5.86 (0.91)	N/A	N/A	>.05
Goal	N/A	N/A	5.08 (0.92)	N/A	N/A	5.63 (0.86)	N/A	N/A	.04[f]
Berger et al (2018) [30]									
Total	3.48 (0.88)	N/A	N/A	3.64 (0.59)	N/A	N/A	N/A	N/A	>.03[g]
Task	3.27 (0.89)	N/A	N/A	3.42 (0.63)	N/A	N/A	N/A	N/A	>.03[g]
Bond	3.60 (1.00)	N/A	N/A	3.98 (0.64)	N/A	N/A	N/A	N/A	>.03[g]
Goal	3.55 (0.99)	N/A	N/A	3.52 (0.74)	N/A	N/A	N/A	N/A	>.03[g]
Lindner et al (2014) [31]									
Total	N/A	N/A	N/A	N/A	N/A	N/A	58.37 (10.55)	N/A	.6
Ly et al (2015) [32]									
Total	65.7 (11.3)	63.5 (9.6)	N/A	N/A	N/A	N/A	N/A	N/A	.75-.37
Meyer et al (2015) [33]								N/A	
Total	N/A	N/A	N/A	N/A	71.0%	N/A	N/A	N/A	N/A
Stiles-Shields et al (2014) [35]									
Total	49.9 (7.57)	N/A	N/A	N/A	N/A	N/A	N/A	49.7 (7.45)	.78
Task	23.3 (4.26)	N/A	N/A	N/A	N/A	N/A	N/A	23.4 (4.15)	.86
Bond	21.9 (5.21)	N/A	N/A	N/A	N/A	N/A	N/A	22.0 (5.13)	.76
Goal	16.5 (2.48)	N/A	N/A	N/A	N/A	N/A	N/A	15.9 (2.44)	.053

[a] Means and standard deviations are displayed, except for the study by Meyer et al, where the percentage of participants rating the alliance as positive is displayed.

[b] F2F: regular face-to-face psychotherapy.

[c] TAU: treatment as usual.

[d] P values based on t tests (Andersson et al and Ly et al), Mann-Whitney U tests (Berger et al), repeated measures analysis of variance (Lindner et al), and least square estimated means (Stiles-Shields et al).

[e] N/A: not applicable.

[f] Statistically significant.

[g] Not statistically significant after Bonferroni correction.

Relationship Between Therapeutic Alliance and Treatment Outcome

Five studies [29,30,32,33,35] investigated the relationship between patients' alliance ratings and treatment outcome (Table 5).

Berger et al [30] found a statistically significant positive association between patients' alliance ratings at 6 weeks and treatment outcome in the face-to-face psychotherapy group ($P<.05$) but not in the combined treatment group. They further showed that residual gain scores of depression were predicted by the patients' alliance rating at 12 weeks in both the regular face-to-face therapy group ($P<.05$) and combined group ($P<.01$). The therapists' ratings at 6 weeks showed no significant association with treatment outcome for either the regular ($P=.61$) or combined treatment ($P=.08$). The therapists' ratings at 12 weeks were significantly associated with treatment outcome in the combined group ($P<.05$) but not in the regular face-to-face therapy group ($P=.60$).

Ly et al [32] reported a significant positive association between patients' working alliance scores and changes in treatment outcome for the blended treatment (*P*<.05) but not for the regular face-to-face psychotherapy.

Meyer et al [33] found a positive correlation between the patients' HAQ-11 scores after 3 weeks of treatment with the internet program plus treatment as usual and symptom reduction. This correlation remained significant (*P*<.02) when controlling for early symptom change.

Anderson et al [29] and Stiles-Shields et al [35] found no relevant relationship between patients' alliance ratings and treatment outcome.

Table 5. Relationship between patients' therapeutic alliance ratings and treatment outcome.

Study, authors (publication year)	F2F[a]	F2F + smartphone	Internet program	Internet program + F2F	Internet program + TAU[b]	Email	Internet program + email/phone	Telephone
Andersson et al (2012) [29]	N/A[c]	N/A	0[d]	N/A	N/A	N/A	N/A	N/A
Berger et al (2018) [30]	+[e]	N/A	N/A	0	N/A	N/A	N/A	N/A
Ly et al (2015) [32]	0	+	N/A	N/A	N/A	N/A	N/A	N/A
Meyer et al (2015) [33]	N/A	N/A	N/A	N/A	+	N/A	N/A	N/A
Stiles-Shields et al (2014) [35]	0	N/A	N/A	N/A	N/A	N/A	N/A	0

[a] F2F: regular face-to-face psychotherapy.

[b] TAU: treatment as usual.

[c] N/A: not applicable.

[d] No significant relationship between therapeutic alliance and treatment outcome.

[e] Statistically significant positive relationship between therapeutic alliance and treatment outcome. Significance based on Spearman correlation (Berger et al, *P*<.01) and mixed effects models (Ly et al, *P*=.00-.05; Meyer et al, *P*<.01).

Discussion

Principal Findings

This study investigated if and how therapeutic alliance ratings were considered in RCTs on different technology-based psychological interventions for depression. Out of the 98 studies included in the TIDECA review, 8 (9.5%) investigated therapeutic alliance. A previous review on e-therapy for different mental health diagnoses found that 1.3% of the included studies considered measures of therapeutic alliance [20]. Notably, the studies included in that review were published between 2002 and 2010, whereas the studies in our review were published between 2012 and 2019. This result may suggest an increased interest in the subject matter; however, it is difficult to directly compare the numbers, since the inclusion criteria of both studies were not identical. Our results showed that the assessment of therapeutic alliance was based on either the WAI or HAQ and that all studies used the patients' ratings, whereas only three studies reported the therapists' ratings as well. Furthermore, this study analyzed therapeutic alliance ratings across different interventions considering factors (eg, the degree of therapeutic guidance) that may affect therapeutic alliance, as well as the relationship between therapeutic alliance and treatment outcome.

It was shown that in a face-to-face psychotherapy setting, the therapeutic relationship was not rated differently when adding an internet-based program as an adjunctive treatment tool. Further, a setting with reduced (four instead of 10) face-to-face sessions and a supportive smartphone app also showed no relevant differences compared with a control setting (10 face-to-face sessions and no smartphone app) regarding therapeutic alliance. Replacing all face-to-face sessions with telephone therapy revealed no relevant difference in alliance ratings. This is especially notable as telephone communication eliminates all visual cues.

In guided internet-based programs, no difference regarding therapeutic alliance was found between guidance via email and guidance via telephone. Thus, in this specific setting, vocal cues and synchronous communication, as provided by the telephone, did not result in other therapeutic alliance ratings compared with asynchronous communication and missing vocal cues. Further, guided internet programs showed alliance ratings similar to those for individualized email therapy.

Overall, therapeutic alliance was rated positively, regardless of the type of technology applied. Notably, this also applied to a setting with no therapist contact at all [33], although it is unclear what exactly is reflected by alliance measures between a person and a program.

Our results are based on one study for each treatment comparison; thus, conclusions need to be considered with caution. Further, therapeutic alliance was assessed at different time points, ranging from 2 to 6 weeks after baseline, which adds heterogeneity to the sample. Thus, different alliance ratings may relate to how the relationship was perceived at a specific time point during the treatment, rather than reflecting differences between the interventions themselves. With this in mind, our results suggest that in the context of TBIs for depression, visual and vocal cues, synchronous communication, and physical presence of the therapist are not requirements for developing good therapeutic alliance. This finding is especially interesting in light of the idea that nonverbal behavior is a key factor of relationship formation between patients and therapists [42]. Possibly, there are other factors that compensate for the lack of nonverbal cues, such as more flexibility when accessing therapeutic modules in web-based programs or smartphone apps. Additionally, there are limited studies about the influence of nonverbal cues on the therapeutic relationship [43]; thus, the relevance of these cues for building an adequate alliance is unclear.

We found mixed results concerning the relationship between therapeutic alliance and treatment outcome (eg, alliance ratings in face-to-face therapy were closely related to the treatment outcome in one study but not in another). It has been argued that alliance ratings are not directly related to treatment outcome, but rather represent a third variable, such as early improvement and treatment motivation [44]. This notion was not confirmed by the results of the study by Meyer et al [33], which showed that early alliance ratings predicted treatment outcome, even when controlling for early symptom change.

Strengths and Limitations

Our review was conducted in line with the Cochrane guidelines, and studies were selected according to prespecified criteria, which were previously published in the study protocol, reflecting high methodological standards. The strict application of inclusion and exclusion criteria reduced the overall number of studies considered in this review. All studies were conducted in Western countries, mainly Europe (Sweden and Germany), and one study was conducted in the United States. These countries share similar communication patterns (eg, relying highly on semantic meaning rather than contextual and nonverbal cues) [45]. Thus, it is unclear whether our results can be extended to non-Western countries, especially in the context of TBIs, where visual and vocal cues are eliminated (eg, email and web-based programs).

There are some additional considerations when interpreting our results. First, the majority of the studies relied upon the WAI for the assessment of therapeutic alliance. As recently pointed out [46], this instrument has not been developed and tested for use in technology-based interventions and may need some adaptions to identify setting-specific influences. This may apply to the specific wording of the instrument (eg, *program* instead of *therapist*), as well as the content of the questions (eg, the goals of the treatment are usually not discussed in a self-help program). Thus, it is possible that future studies utilizing a tool that is specifically developed and validated in this context may show divergent results.

Second, not all studies specifically reported whether any additional contact between therapists and patients occurred outside of the treatment setting. For example, if a patient is assigned to an email therapy group but meets the therapist during the recruitment or initial assessment process, additional exposure to visual and vocal cues could influence the formation of a therapeutic bond. Thus, we encourage future studies to take this point into consideration by reporting all contacts between therapists and patients.

Third, most patients were recruited via advertisements. Previous research has shown that patients recruited from nonclinical settings have a more positive attitude toward internet interventions for the treatment of depression than patients recruited from clinical settings [47]. Further, therapists that agreed to participate in the studies may have been subject to a similar selection bias. Thus, the results may not be transferrable to clinical practice, where more skeptical individuals (concerning TBIs) are present.

Fourth, only three out of eight studies took the therapists' perception of therapeutic alliance into consideration. Previous studies have shown that therapists and patients may judge the alliance differently [13]; thus, it would be favorable to include both measures in future studies.

Implications for Clinical Practice and Future Research

As the field may continue to expand and more therapists may consider technology-based treatment options, it is important to further investigate on what basis the alliance is formed in settings that do not allow for additional cues, such as facial expression and tone. For example, in an RCT with several arms, each eliminating a different component of face-to-face communication (eg, visual or auditory cues), researchers could investigate if and how therapeutic alliance is affected. Importantly, these studies should include both the therapists' and patients' perspectives. Further, it will be important to replicate the finding that early alliance ratings may predict treatment outcome independently from early symptom change. Such studies will require more frequent measurements of therapeutic alliance, symptom change, and alternative factors that could relate to treatment outcome, such as pretreatment motivation, in order to establish the relation of these aspects. Moreover, research needs to address how personal preferences and attitudes toward TBIs may interact with the formation of a therapeutic alliance, both from the patient's and therapist's perspective. Finally, it remains unclear what exactly is measured in settings without any therapist contact at all. Possibly, the alliance between a person and a technology-based program reflects less of the emotional aspect of therapeutic alliance and is more related to how the program matches an individual's goals and expectations of the tasks required.

Conclusion

This review shows that studies on therapeutic alliance in TBIs for the treatment of depression are still limited, especially regarding therapists' alliance ratings. Taking into account different degrees of therapeutic guidance, qualifications of therapists, modes of delivery, and types of technologies, the results of this review suggest that a positive therapeutic alliance can be established in TBIs for people with depression.

Acknowledgments

The Federal Ministry of Education and Research (in German: Bundesministerium für Bildung und Forschung; BMBF) funded the systematic review on the comparative effectiveness of *T*echnology-based *I*nterventions in different steps of *De*pression *Ca*re (TIDECA; funding code: 01KG1705, funding period: August 2017 to October 2019). The funding institution had no role in the design of this study and no role during its execution, the analysis and interpretation of data, or the decision to submit the results. The review on therapeutic alliance was conducted as part of a doctoral degree (medicine) and did not receive any additional funding. We thank Mareike Dreier (MD) for her assistance in screening the full-text publications for inclusion.

Conflicts of Interest

None declared.

References

1. World Health Organization. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates URL: https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/ [accessed 2019-08-29]
2. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, CDBE2010 Study Group. Cost of disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Oct;21(10):718-779. [doi: 10.1016/j.euroneuro.2011.08.008] [Medline: 21924589]
3. Cuijpers P, Van Straten A, Warmerdam L, Smits N. Characteristics of effective psychological treatments of depression: a metaregression analysis. Psychother Res 2008 Mar;18(2):225-236. [doi: 10.1080/10503300701442027] [Medline: 18815968]
4. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
5. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
6. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
7. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
8. Mohr DC, Vella L, Hart S, Heckman T, Simon G. The Effect of Telephone-Administered Psychotherapy on Symptoms of Depression and Attrition: A Meta-Analysis. Clin Psychol (New York) 2008;15(3):243-253 [FREE Full text] [doi: 10.1111/j.1468-2850.2008.00134.x] [Medline: 21369344]
9. Twomey C, O'Reilly G, Meyer B. Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: A meta-analysis. Psychiatry Res 2017 Oct;256:371-377. [doi: 10.1016/j.psychres.2017.06.081] [Medline: 28686935]
10. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health. European Psychologist 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
11. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychological Medicine 2014 Nov;44(15):3137-3150. [doi: 10.1017/s0033291714000245] [Medline: 25065947]
12. Townsend L, Gearing RE, Polyanskaya O. Influence of health beliefs and stigma on choosing internet support groups over formal mental health services. Psychiatr Serv 2012 Apr;63(4):370-376. [doi: 10.1176/appi.ps.201100196] [Medline: 22337006]
13. Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. Psychotherapy (Chic) 2011 Mar;48(1):9-16. [doi: 10.1037/a0022186] [Medline: 21401269]
14. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research & Practice 1979;16(3):252-260. [doi: 10.1037/h0085885]
15. Berger T. The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
16. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology 1989 Apr;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
17. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
18. Falkenström F, Ekeblad A, Holmqvist R. Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. J Consult Clin Psychol 2016 Aug;84(8):738-751. [doi: 10.1037/ccp0000119] [Medline: 27213493]
19. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol 2000 Jun;68(3):438-450. [Medline: 10883561]

20. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 02;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]
21. Preschl B, Maercker A, Wagner B. The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. BMC Psychiatry 2011 Dec 6;11(1):1-10. [doi: 10.1186/1471-244x-11-189]
22. Beckner V, Vella L, Howard I, Mohr DC. Alliance in two telephone-administered treatments: relationship with depression and health outcomes. J Consult Clin Psychol 2007 Jun;75(3):508-512. [doi: 10.1037/0022-006X.75.3.508] [Medline: 17563167]
23. Köhnen M, Kriston L, Härter M, Dirmaier J, Liebherz S. Rationale and design of a systematic review: effectiveness and acceptance of technology-based psychological interventions in different clinical phases of depression management. BMJ Open 2019 Mar 27;9(3):e028042 [FREE Full text] [doi: 10.1136/bmjopen-2018-028042] [Medline: 30918040]
24. Higgins JP, Green S. Cochrane Training. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] URL: https://training.cochrane.org/handbook/archive/v5.1/ [accessed 2019-08-29]
25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
26. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997 Sep 01;127(5):380-387. [doi: 10.7326/0003-4819-127-5-199709010-00008] [Medline: 9273830]
27. Eaton TT, Abeles N, Gutfreund MJ. Therapeutic alliance and outcome: Impact of treatment length and pretreatment symptomatology. Psychotherapy: Theory, Research, Practice, Training 1988;25(4):536-542. [doi: 10.1037/h0085379]
28. Machmutow K, Meister R, Jansen A, Kriston L, Watzke B, Härter MC, et al. Comparative effectiveness of continuation and maintenance treatments for persistent depressive disorder in adults. Cochrane Database Syst Rev 2019 May 20;5:CD012855 [FREE Full text] [doi: 10.1002/14651858.CD012855.pub2] [Medline: 31106850]
29. Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, et al. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. Behav Res Ther 2012 Sep;50(9):544-550. [doi: 10.1016/j.brat.2012.05.003] [Medline: 22728647]
30. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial. J Affect Disord 2018 Feb;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
31. Lindner P, Olsson EL, Johnsson A, Dahlin M, Andersson G, Carlbring P. The impact of telephone versus e-mail therapist guidance on treatment outcomes, therapeutic alliance and treatment engagement in Internet-delivered CBT for depression: A randomised pilot trial. Internet Interventions 2014 Oct;1(4):182-187. [doi: 10.1016/j.invent.2014.09.001]
32. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-Supported versus Full Behavioural Activation for Depression: A Randomised Controlled Trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
33. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. Internet Interventions 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
34. Steinmann M, Heddaeus D, Liebherz S, Daubmann A, Härter M, Watzke B. Effectiveness of Telephone-Administered Cognitive-Behavioral Psychotherapy for Depression With Versus Without Additional Letters: A Randomized Controlled Trial. Telemed J E Health 2020 Mar;26(3):347-353. [doi: 10.1089/tmj.2018.0311] [Medline: 31013466]
35. Stiles-Shields C, Kwasny MJ, Cai X, Mohr DC. Therapeutic alliance in face-to-face and telephone-administered cognitive behavioral therapy. J Consult Clin Psychol 2014 Apr;82(2):349-354 [FREE Full text] [doi: 10.1037/a0035554] [Medline: 24447003]
36. Zwerenz R, Becker J, Knickenberg RJ, Siepmann M, Hagen K, Beutel ME. Online Self-Help as an Add-On to Inpatient Psychotherapy: Efficacy of a New Blended Treatment Approach. Psychother Psychosom 2017;86(6):341-350. [doi: 10.1159/000481177] [Medline: 29131090]
37. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
38. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. Psychological Assessment: A Journal of Consulting and Clinical Psychology 1989 Sep;1(3):207-210. [doi: 10.1037/1040-3590.1.3.207]
39. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychotherapy Research 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
40. Bassler M, Potratz B, Krauthauser H. Der Helping Alliance Questionnaire (HAQ) von Luborsky. Psychotherapeut 1995;40:23-32. [doi: 10.1037/t16583-000]
41. Alexander LB, Luborsky L. The Penn Helping Alliance Scales. In: Greenberg LS, Pinsof WM, editors. Guilford clinical psychology and psychotherapy series. The psychotherapeutic process: A research handbook. New York: Guilford Press; 1986:325-366.

42. Hall JA, Harrigan JA, Rosenthal R. Nonverbal behavior in clinician—patient interaction. Applied and Preventive Psychology 1995 Dec;4(1):21-37. [doi: 10.1016/s0962-1849(05)80049-6]
43. Ramseyer F, Tschacher W. Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. J Consult Clin Psychol 2011 Jun;79(3):284-295. [doi: 10.1037/a0023419] [Medline: 21639608]
44. Strunk DR, Cooper AA, Ryan ET, DeRubeis RJ, Hollon SD. The process of change in cognitive therapy for depression when combined with antidepressant medication: Predictors of early intersession symptom gains. J Consult Clin Psychol 2012 Oct;80(5):730-738 [FREE Full text] [doi: 10.1037/a0029281] [Medline: 22774791]
45. Qureshi A, Collazos F. The intercultural and interracial therapeutic relationship: challenges and recommendations. Int Rev Psychiatry 2011;23(1):10-19. [doi: 10.3109/09540261.2010.544643] [Medline: 21338293]
46. Gómez Penedo JM, Berger T, Grosse Holtforth M, Krieger T, Schröder J, Hohagen F, et al. The Working Alliance Inventory for guided Internet interventions (WAI-I). J Clin Psychol 2019 Jun 25:1-32. [doi: 10.1002/jclp.22823] [Medline: 31240727]
47. Schröder J, Berger T, Meyer B, Lutz W, Hautzinger M, Späth C, et al. Attitudes Towards Internet Interventions Among Psychotherapists and Individuals with Mild to Moderate Depression Symptoms. Cogn Ther Res 2017 Apr 22;41(5):745-756. [doi: 10.1007/s10608-017-9850-0]

Abbreviations

HAQ: Helping Alliance Questionnaire
HAQ-11: 11-item Helping Alliance Questionnaire
RCT: randomized controlled trial
TAU: treatment as usual
TBI: technology-based intervention
TIDECA: Technology-based Interventions in different steps of Depression Care
WAI: Working Alliance Inventory
WAI-S: Working Alliance Inventory short form version
WAI-SR: Working Alliance Inventory short form-revised version

---

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by H Baumeister, M Eggeling; comments to author 23.12.19; revised version received 15.02.20; accepted 02.03.20; published 11.06.20.

Please cite as:
Wehmann E, Köhnen M, Härter M, Liebherz S
Therapeutic Alliance in Technology-Based Interventions for the Treatment of Depression: Systematic Review
J Med Internet Res 2020;22(6):e17195
URL: http://www.jmir.org/2020/6/e17195/
doi:10.2196/17195
PMID:32525484

Original Paper

Two Decades of Research Using Taiwan’s National Health Insurance Claims Data: Bibliometric and Text Mining Analysis on PubMed

Sheng-Feng Sung[1,2*], MD, MSc; Cheng-Yang Hsieh[3,4*], MD, PhD; Ya-Han Hu[5,6,7], PhD

[1]Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chiayi Christian Hospital, Chiayi City, Taiwan

[2]Department of Information Management, Institute of Healthcare Information Management, National Chung Cheng University, Chiayi County, Taiwan

[3]Department of Neurology, Tainan Sin Lau Hospital, Tainan, Taiwan

[4]School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[5]Department of Information Management, National Central University, Taoyuan City, Taiwan

[6]Center for Innovative Research on Aging Society, National Chung Cheng University, Chiayi County, Taiwan

[7]MOST AI Biomedical Research Center, National Cheng Kung University, Tainan, Taiwan

*these authors contributed equally

Corresponding Author:
Ya-Han Hu, PhD
Department of Information Management
National Central University
No 300, Zhongda Rd, Zhongli District
Taoyuan City, 32001
Taiwan
Phone: 886 3 4227151 ext 66560
Fax: 886 3 4254604
Email: yhhu@mgt.ncu.edu.tw

Abstract

Background: Studies using Taiwan’s National Health Insurance (NHI) claims data have expanded rapidly both in quantity and quality during the first decade following the first study published in 2000. However, some of these studies were criticized for being merely data-dredging studies rather than hypothesis-driven. In addition, the use of claims data without the explicit authorization from individual patients has incurred litigation.

Objective: This study aimed to investigate whether the research output during the second decade after the release of the NHI claims database continues growing, to explore how the emergence of open access mega journals (OAMJs) and lawsuit against the use of this database affect the research topics and publication volume and to discuss the underlying reasons.

Methods: PubMed was used to locate publications based on NHI claims data between 1996 and 2017. Concept extraction using MetaMap was employed to mine research topics from article titles. Research trends were analyzed from various aspects, including publication amount, journals, research topics and types, and cooperation between authors.

Results: A total of 4473 articles were identified. A rapid growth in publications was witnessed from 2000 to 2015, followed by a plateau. Diabetes, stroke, and dementia were the top 3 most popular research topics whereas statin therapy, metformin, and Chinese herbal medicine were the most investigated interventions. Approximately one-third of the articles were published in open access journals. Studies with two or more medical conditions, but without any intervention, were the most common study type. Studies of this type tended to be contributed by prolific authors and published in OAMJs.

Conclusions: The growth in publication volume during the second decade after the release of the NHI claims database was different from that during the first decade. OAMJs appeared to provide fertile soil for the rapid growth of research based on NHI claims data, in particular for those studies with two or medical conditions in the article title. A halt in the growth of publication volume was observed after the use of NHI claims data for research purposes had been restricted in response to legal controversy. More efforts are needed to improve the impact of knowledge gained from NHI claims data on medical decisions and policy making.

(J Med Internet Res 2020;22(6):e18457) doi:10.2196/18457

KEYWORDS

administrative claims data; bibliometric analysis; National Health Insurance; text mining; open access journals; PubMed

Introduction

Health care administrative data, also known as administrative claims data, [1], are derived from claims for reimbursement for routine health care services. They are relatively inexpensive to procure and, in general, readily available in electronic format [2]. Therefore, they are widely used for medical and public health research [1,3,4]. Such claims-based studies may cover a variety of research types, such as disease surveillance, health service utilization, validity analysis, and association between exposure and health outcomes [5].

Taiwan's National Health Insurance Research Database (NHIRD), one of the largest health care administrative databases in the world, has provided a great opportunity for researchers to perform population-based studies [6]. Since 1995, residents in Taiwan have enjoyed a universal single-payer health care system operated by the National Health Insurance (NHI). The program covered virtually all of Taiwan's population (99.5%) by 2010. The coverage of the whole population in the database has the advantages of an enormous sample size and lack of participation bias [6]. In 2000, Taiwan's National Health Research Institutes compiled NHI claims data into the NHIRD and made it publicly available to the academic community. Low-cost updates of the NHIRD are possible because its data collection is carried out routinely for purposes inherent in the process of medical care and insurance reimbursement [6]. The public release of this large population-based database enables collaboration and knowledge sharing among researchers and boosts scientific production [6,7]. A bibliometric study conducted by Chen et al [8] to investigate the trend of using the NHIRD as research material found that the number of these studies had increased significantly from 2000 to 2009, with an average annual growth rate of 45.8%.

While more and more authors have successfully published studies using NHI claims data as the data source, some of these studies were criticized for being merely data-dredging studies rather than hypothesis-driven [9]. For example, as described by Hampson et al [9], increased risks of 5 different medical conditions following carbon monoxide poisoning were reported in 5 individual articles using the same research model. In addition, the emergence of open access mega journals (OAMJs) since 2006 [10] may have played a role. OAMJs, as a new business model for publication, are characterized by a large publication volume and broad disciplinary scope. They generally accept research articles as long as the requirements of "scientific soundness" are met, regardless of their contributions to a research field or academic interest [10]. It is believed, justified or not, that Journal Citation Reports (JCR)-indexed OAMJs are especially attractive to some researchers in Taiwan [10,11].

On the other hand, with the widespread use of the NHIRD in research, several human rights groups protested the use of claims data without the explicit authorization from individual patients and launched a lawsuit against the NHI Administration in 2012 [12,13]. As a result, the National Health Research Institutes stopped accepting applications for the NHIRD from researchers at the end of 2015 and terminated the NHIRD service in mid-2016 [14]. Thereafter, researchers can only access NHI claims data on site at the Data Science Center or via a virtual private network connection at its branch offices [15].

Studies based on the NHIRD have expanded substantially in both quantity and quality during the period of 2000 to 2009 [8]. However, whether the recently burgeoning OAMJs or lawsuit would affect the research output from NHI claims data remained undetermined. Since Asian countries like Japan and South Korea are also developing their own nationwide claim databases [16], the lessons regarding how Taiwan's NHI claims database contributed to academic production may help researchers in other countries to develop their own policy and strategy on how to use claims databases in research. Accordingly, we formulated the following research questions: (1) With the widespread use of NHI claims data as research material, does the number of research articles in the second decade keep growing at a similar pace to that observed in the first decade since the release of NHI claims data? What is the effect of the lawsuit against the secondary use of health insurance data for research on research output? (2) What are the most common research topics of articles using NHI claims data? Do the types of research topics correlate with the volume of research output? (3) Which journals published the most articles using NHI claims data? (4) What is the role of open access journals in the proliferation of research output? (5) Who are the most prolific authors and research groups?

Methods

Data Source

This study used PubMed to locate publications that may have used NHI claims data as the primary data source because PubMed is the most widely used database for searching medical literature. Articles published in English that entered PubMed between Jan 1, 1996 and Dec 31, 2017 with "journal article" as their publication type were included. Following the work by Chen et al [8], a broad search strategy was employed to permit inclusion of as many articles as possible. Articles had to mention "Taiwan" in any of the textual fields including title, abstract, medical subject heading (MeSH) index terms, and author's affiliation address and meet any of the following criteria: (1) indexed with the MeSH term "insurance, health" or "national health programs;" (2) either "nationwide" or "population" in the title field; (3) any of the following terms appearing in the title or abstract: "health insurance," "national insurance," "claims data*," "claim data*," "insurance claim*," "insurance data*," "administrative data*," "nationwide data*," "national data*," "NHIRD," "LHID," "NHI," and "BNHI." The asterisk (*) is the truncation symbol used by PubMed that indicates to find all terms that begin with the string preceding the asterisk.

Articles classified under the categories of comment, letter, editorial, or review and those published without an abstract were excluded. The search was done on June 24, 2018 and resulted in 5105 articles for further evaluation (See Multimedia Appendix 1).

Because journals may change their titles or even merge, this study always adopted the last title of a journal. JCR Science Edition and Social Sciences Edition (Clarivate Analytics, 2018) was used to retrieve 2017 Journal Impact Factors and journal categories. Journals were classified from Q1 to Q4 according to the impact factor quartiles in the specific journal category, where Q1 journals stand for journals with higher impact factors. Journals not indexed by the JCR were classified as other. Open access journals were identified through the Directory of Open Access Journals. OAMJs were defined as described in a previous study [10].

Ascertainment of Studies

All the articles were downloaded from PubMed and preprocessed using the "easyPubMed" package in R. Because the list of potential articles was quite lengthy, several heuristic rules were applied to determine whether an article used NHI claims data as the primary data source. Basically, regular expression pattern matching was used to detect the mentioning of using NHI claims data in article abstracts and adjusted the matching patterns by trial and error. This study finally found two inclusion rules and one exclusion rule. The rules are shown in Textbox 1. These rules identified 3059 articles as using NHI claims data and achieved 100% precision (positive predictive value) by manually examining a random sample of 500 articles.

The remaining 2046 articles were reviewed by the first and second authors. Each author independently classified an article as "using NHI claims data," "not using NHI claims data," or "using data from undetermined source" by examining its abstract or full text, when necessary. This process achieved an agreement of 99.1% (kappa=0.978), and discrepancies (19 articles) were resolved by consensus. Among them, 632 articles were considered not using NHI claims data and thus excluded. In the end, a total of 4473 articles were included in this study (Multimedia Appendix 1).

Textbox 1. Heuristic rules used to determine whether an article used National Health Insurance claims data.

Inclusion rules

1. (from|data|study|using|cohort|used|based|patients|identified|population|obtained|

claim|conducted|retrieved|collected|selected|analyz)[[:print:]]{1,20}(National Health Insurance|NHI|Longitudinal Health Insurance|insurance claims|Registry for Catastrophic Illness Patient)[[:print:]]{1,20}(claim|data|file)")

2. (National Health Insurance|NHI|Longitudinal Health Insurance|insurance claims|Registry for Catastrophic Illness Patients)[[:print;]]{1,20}(claim|data|file)[[:print:]]{1,20}(from|

used|patients|identified)

Exclusion rule

1. (Korea)[[:print:]]{1,20}(National Health Insurance|NHI)

Text Mining

This study used MetaMap as the tool to mine medical entities, such as symptoms, clinical findings, diseases, and medications, from article titles. MetaMap is a natural language processing tool developed by the National Library of Medicine. It analyzes input text through tokenization, sentence boundary determination, part-of-speech tagging, and parsing and generates variants of resulting phrases or words [17]. By evaluating measures of centrality, variation, coverage, and cohesiveness, MetaMap locates each matched medical entity in the Unified Medical Language System (UMLS) Metathesaurus, assigns it a semantic type, and returns a concept unique identifier and score between 0 and 1000, with a higher value representing a closer match [18]. This study used UMLS Metathesaurus version 2016AA.

Because an article title typically indicates what the article is about, this study attempted to mine knowledge from article titles. Although MeSH terms are also used in PubMed to describe what an article is about, we analyzed MetaMap-derived concepts instead of MeSH concepts for two reasons. First, we focused on the article title, but the MeSH concepts were determined by examining the whole article; second, the UMLS Metathesaurus contains far more medical concepts than the MeSH vocabulary. This study focused on two categories of medical entities: (1) medical conditions including diseases, symptoms and signs, and findings and (2) interventions, including medications, procedures, and surgery. Specifically, medical entities were categorized based on their UMLS semantic types (see Multimedia Appendix 2). Because MetaMap may generate multiple UMLS semantic types and concepts from the same phrase [18], this study relied on the order output by MetaMap and accepted only the first returned semantic type and concept. Furthermore, MetaMap might retrieve general terms that are not the medical entities the researchers were interested in, such as "adopt," "75+ years," and "ambulatory visit". Therefore, this study collected these terms in a stop word list by manually reviewing an aggregate list of the concepts returned by MetaMap. The terms in the stop word list were then excluded from analysis.

Researchers may have different preferences for research topics when using administrative databases as the primary data source [5]. Descriptive studies may investigate only one medical condition whereas analytical studies generally focus on the association between an exposure (either a medical condition or an intervention) and an outcome (a medical condition) [19]. Some studies that used NHI claims data were considered to replicate the same research model of examining the association between two medical conditions [9]. Motivated by this criticism, this study classified articles into 4 study types based on medical entities mentioned in the title as follows: (1) with ≥1 intervention regardless of the number of medical conditions, (2) with ≥2 medical conditions but without any intervention, (3) with only 1 medical condition but without any intervention, and (4) others. Each article was assigned to only 1 of the 4 study types. For example, in the article title "Tamoxifen and the risk of Parkinson's disease in female patients with breast cancer in Asian people," tamoxifen is an intervention whereas Parkinson's disease and breast cancer are both medical conditions. This article is classified as the type with ≥1 intervention. An article with the title "Increased risk of stroke in patients with chronic kidney disease after recurrent hypoglycemia" mentions three medical conditions and is classified as the type with ≥2 medical conditions but without any intervention.

In order to offer background statistics, PubMed was queried to identify the number of articles published in English that entered PubMed between 2000 and 2017 with "journal article" as their publication type. The number of articles published between 2000 and 2017 in the top 20 journals that have published the most studies using NHI claims data were also obtained. Furthermore, the titles of the articles in PubMed and the top 20 journals were screened for presence of the top 10 medical conditions and top 10 interventions that were the most prevalent among studies using NHI claims data. The percentage of each medical condition or intervention among published articles was calculated.

Statistical Analysis and Data Visualization

Categorical variables are reported as counts (percentages). Comparisons between groups used chi-square tests. Trends in continuous variables were assessed using the Cuzick test. Trends in categorical outcomes were evaluated using the Cochran-Armitage trend test for binomial proportions and the multinomial Cochran-Armitage Trend Test implemented in the R package "multiCA" [20] for multinomial proportions. Social network analysis was applied to explore the cooperation between authors and the relationships between medical entities. The "igraph" package in R was used to produce the network graph. Each node represents an author or a medical entity. The nodes are joined by weighted links, in which the width of a link indicates the frequency of relationships between two nodes. The size of a node is proportional to the weighted degree centrality of the node, which is computed by summing the weights of links to the node.

Statistical analyses and visualizations were performed using Stata 15.1 (StataCorp, College Station, Texas) and R version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria). Two-tailed *P* values were considered statistically significant at <.05.

Results

Trends in Research Output

Since the first article appeared in 2000, the number of publications grew tremendously until 2015, when the publication output seemed to reach a plateau (Figure 1). In contrast, the number of publications in PubMed and the top 20 journals that have published the most studies using NHI claims data continued to increase after 2015 (Figure 1). By 2015, the average annual growth rate of published articles was 77.4%, with a doubling time of 1.7 years. Almost all the articles (97.0%) were published in journals indexed in the JCR Science Edition or Social Sciences Edition. Among the articles indexed in the JCR, 46.4%, 38.3%, 11.5%, and 3.8% were published in Q1, Q2, Q3, and Q4 journals, respectively. Figure 1 illustrates the distribution of articles across the four quartiles of journals in the JCR each year. Table 1 gives the characteristics of articles, authors, and journals across 3 time periods and the results of the trend analysis by year. Based on the affiliation of the first author, we determined that researchers in nonhospital institutions published most of the articles initially, which were gradually outnumbered by those produced by hospital researchers in recent years. Articles were increasingly published in open access journals, in particular OAMJs. Articles were published in an increasing number of journals and were more widely distributed across disciplines.

Figure 1. Annual number of publications between 2000 and 2017 (A) in PubMed, (B) in the top 20 journals that have published the most studies using National Health Insurance (NHI) claims data, and (C) based on NHI claims data, separated by JCR (2017 edition) ranking (Q1, Q2, Q3, Q4). The first year of each major OAMJ is indicated. The release of the National Health Insurance Research Databases (NHIRD) began in 2000 and ended in 2015. The Health and Welfare Database (HWD) was created in 2009 and is still available for research use.

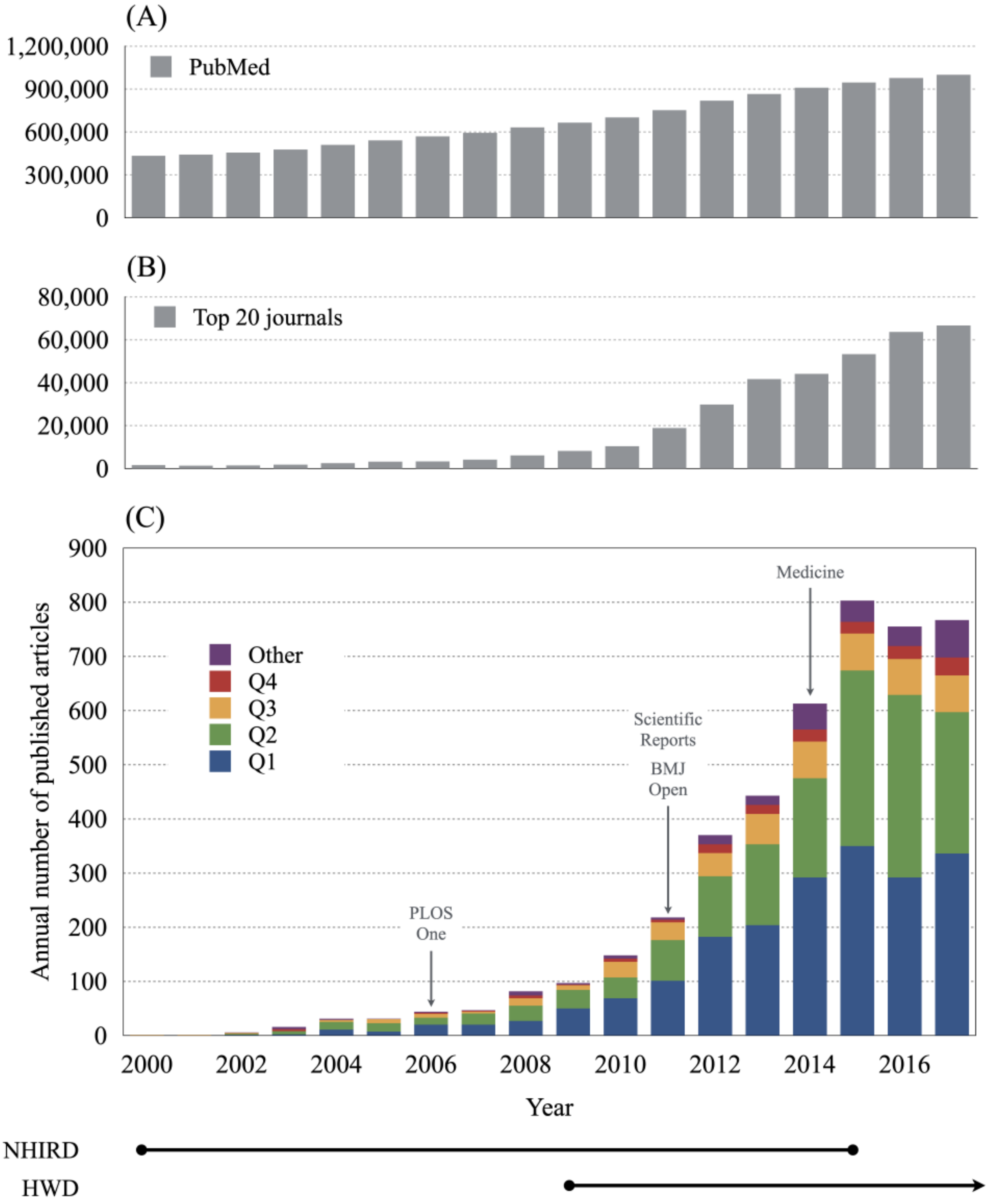


Table 1. Characteristics of and trends for articles and journals.

Articles and journals	Period			Total number	*P* value for trend
	2000–2005	2006–2011	2012–2017		
Articles					
Indexed in PubMed, n	86	636	3751	4473	<.001
Indexed in JCR[a] 2017, n (%)	78 (90.7)	625 (98.3)	3635 (96.9)	4338 (97.0)	.219
First author from hospitals, n (%)	33 (38.4)	303 (47.6)	2479 (66.1)	2815 (62.9)	<.001
First author from abroad	3 (3.5)	20 (3.1)	71 (1.9)	94 (2.1)	.019
Published in OAJs[b], n (%)	16 (18.6)	103 (16.2)	1447 (38.6)	1566 (35.0)	<.001
Published in OAMJs[c], n (%)	0 (0)	8 (1.3)	898 (23.9)	906 (20.3)	<.001
Study types per article title					<.001
With ≥1 intervention, n (%)	29 (33.7)	202 (31.8)	1315 (35.1)	1546 (34.6)	.002
With ≥2 conditions, n (%)	8 (9.3)	132 (20.8)	1435 (38.3)	1575 (35.2)	<.001
With only 1 condition, n (%)	30 (34.9)	210 (33.0)	797 (21.2)	1037 (23.2)	<.001
Others, n (%)	19 (22.1)	92 (14.5)	204 (5.4)	315 (7.0)	<.001
Journals					
Indexed in PubMed, n	59	297	766	841	<.001
Indexed in JCR 2017, n	53	287	727	791	<.001
Journal categories, n	32	56	68	68	<.001

[a] JCR: Journal Citation Reports.

[b] OAJ: open access journal.

[c] OAMJ, open access mega journal.

Research Topics of Articles

A total of 1763 medical entities were retrieved from article titles using MetaMap. Among these, 1132 entities belonged to the category of medical conditions, and 631 belonged to the category of interventions. The most commonly investigated medical conditions were diabetes (n=263), stroke (n=189), and dementia (n=139), whereas the most commonly studied interventions were statin (n=100), metformin (n=40), and Chinese herbal medicine (n=38). The top 50 medical conditions and interventions are listed in Table 2.

Table 3 gives the number and percentage of articles with one of the top 10 medical conditions or the top 10 interventions in the article title during the same period (2000 to 2017). Apparently, the top 10 medical conditions and top 10 interventions were not so prevalent in the titles among articles in PubMed or the top 20 journals.

Figure 2 illustrates the distribution of study types over the years. A significant trend in the proportions of study types was found ($P<.001$). In particular, the number of articles with ≥2 medical conditions in the title grew extensively whereas the number of articles with only one medical condition or without any medical entities in the title decreased considerably (Table 1). Figure 3 shows the most common condition-condition pairs and condition-intervention pairs in article titles. Diabetes, stroke, and dementia remained the top three medical conditions that were extensively studied in association with other medical conditions. They co-occurred with 118, 111, and 80 other medical conditions in article titles, respectively. Among the most studied interventions, statin, metformin, and Chinese herbal medicine co-occurred with 70, 31, and 28 medical conditions in article titles, respectively.

Table 2. Relative frequency of the top 50 medical conditions and interventions mentioned in article titles of studies based on National Health Insurance (NHI) claims data.

Medical conditions and interventions	Number of times mentioned
Medical conditions	
Diabetes	263
Stroke	189
Dementia	139
Type 2 diabetes	132
Cancer	124
End stage renal disease	102
Atrial fibrillation	88
Chronic obstructive pulmonary disease	88
Chronic kidney disease	87
Schizophrenia	82
Ischemic stroke	80
Asthma	79
Depression	74
Rheumatoid arthritis	71
Breast cancer	68
Tuberculosis	65
Parkinson disease	59
Acute myocardial infarction	55
Bipolar disorder	55
Hepatocellular carcinoma	55
Hypertension	55
Osteoporosis	55
Hip fracture	54
Lupus erythematosus, systemic	54
Pneumonia	53
Fracture	51
Attention deficit-hyperactivity disorder	49
Acute pancreatitis	46
Cardiovascular disease	45
Erectile dysfunction	41
Lung cancer	41
Acute coronary syndrome	40
Coronary artery disease	38
Depressive disorder	38
Infection	37
Peripheral arterial disease	37
Prostate cancer	37
Epilepsy	36
Traumatic brain injury	36
Sleep disorder	35

Medical conditions and interventions	Number of times mentioned
Colorectal cancer	34
Migraine	34
Psoriasis	34
Hearing loss, sudden	31
Gout	27
Liver abscess, pyogenic	27
Obstructive sleep apnea	27
Sleep apnea	27
Alzheimer's disease	26
Psychiatric disorder	26
Interventions	
Statin	100
Metformin	40
Chinese herbal medicine	38
Hemodialysis	38
Antidepressant	36
Antipsychotic	35
Proton pump inhibitor	29
Dialysis	27
Nonsteroidal anti-inflammatory drugs	27
Corticosteroid	24
Influenza vaccination	24
Angiotensin-converting enzyme inhibitor	22
Benzodiazepine	21
Zolpidem	19
Antihypertensive agents	18
Dialysis, peritoneal	17
Thiazolidinedione	17
Antidiabetic	16
Angiotensin receptor blockers	14
Reduction	14
Tamoxifen	14
Antibiotic	13
Cholecystectomy	13
Sitagliptin	13
Angiotensin 2 receptor blockers	12
Antiepileptic drug	12
Chemotherapy	12
Selective serotonin reuptake inhibitors	12
Acupuncture	11
Intervention, percutaneous coronary	11
Mechanical ventilation	11
Pioglitazone	11

Medical conditions and interventions	Number of times mentioned
Appendectomy	10
Aspirin	10
Clopidogrel	10
Hormone therapy	10
Interferon	10
Splenectomy	10
Total knee arthroplasty	10
Antiplatelet agents	9
Caesarian section	9
Coronary artery bypass grafting	9
Drug eluting stent	9
Liver transplantation	9
Radiotherapy	9
Resection	9
Alendronate	8
Antiviral	8
Digoxin	8
Hypnotic	8

Table 3. Number and percentage of articles with the corresponding condition or intervention in the article title between 2000 and 2007.

Articles	Studies using NHI[a] claims data (n=4473), n (%)	Articles in top 20 journals n=362,463, n (%)	Articles in PubMed n=12,309,239, n (%)
Articles with the condition in the article title			
Diabetes	263 (5.9)	4195 (1.2)	112,484 (0.9)
Stroke	189 (4.2)	1944 (0.5)	56,781 (0.5)
Dementia	139 (3.1)	664 (0.2)	24,217 (0.2)
Type 2 diabetes	132 (3.0)	1724 (0.5)	38,757 (0.3)
Cancer	124 (2.8)	22,030 (6.1)	503,283 (4.1)
End stage renal disease	102 (2.3)	169 (0.0)	4387 (0.0)
Atrial fibrillation	88 (2.0)	1189 (0.3)	21,867 (0.2)
Chronic obstructive pulmonary disease	88 (2.0)	450 (0.1)	10,356 (0.1)
Chronic kidney disease	87 (1.9)	671 (0.2)	12,617 (0.1)
Schizophrenia	82 (1.8)	1215 (0.3)	33,536 (0.3)
Articles with the intervention in the article title			
Statin	100 (2.2)	319 (0.1)	5130 (0.0)
Metformin	40 (0.9)	372 (0.1)	6408 (0.1)
Chinese herbal medicine	38 (0.8)	158 (0.0)	674 (0.0)
Hemodialysis	38 (0.8)	58 (0.0)	3175 (0.0)
Antidepressant	36 (0.8)	696 (0.2)	7327 (0.1)
Antipsychotic	35 (0.8)	337 (0.1)	5540 (0.0)
Proton pump inhibitor	29 (0.6)	38 (0.0)	1105 (0.0)
Dialysis	27 (0.6)	416 (0.1)	16,301 (0.1)
Nonsteroidal anti-inflammatory drugs	27 (0.6)	2 (0.0)	260 (0.0)
Corticosteroid	24 (0.5)	109 (0.0)	4475 (0.0)

[a] NHI: National Health Insurance.

Figure 2. Distribution of study types across the years.

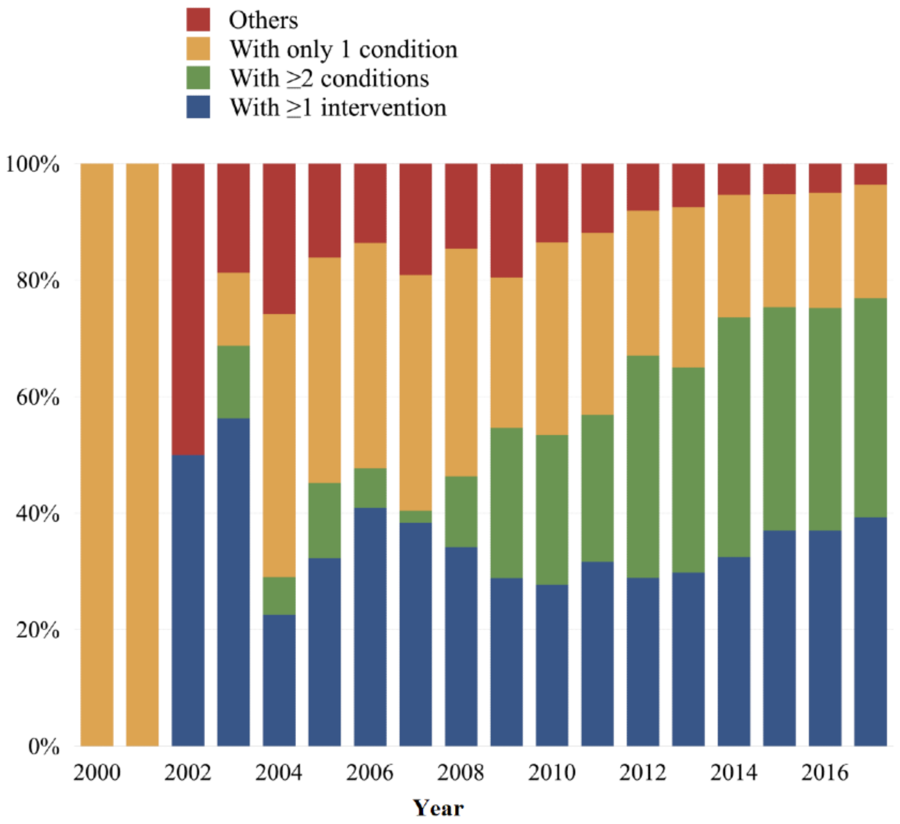
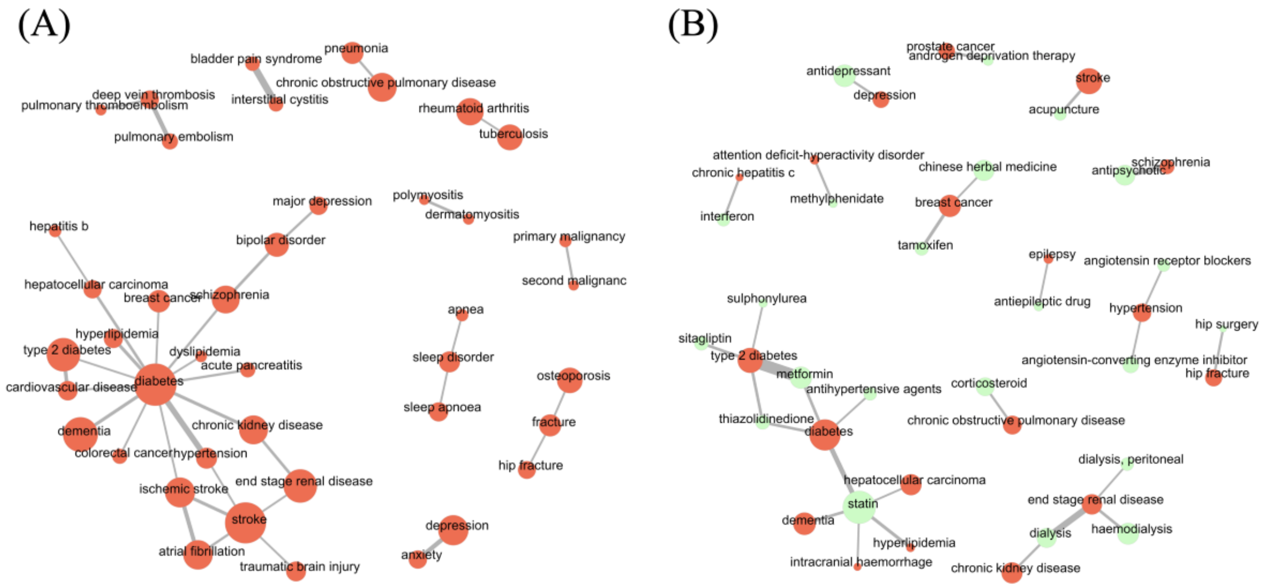


Figure 3. Network graphs displaying the most common (A) condition-condition pairs and (B) condition-intervention pairs.



Scattering of Articles in Journals

Until the end of 2017, 4473 articles were published in 841 journals, with an average of 5.3 articles per journal. Among these journals, 791 were indexed in the JCR Science Edition or Social Sciences Edition and were spread across 68 disciplines. The journals PLOS ONE and Medicine ranked first and second, respectively, in the number of publications and yielded 18.0% (804/4473) of the articles, whereas 333 journals published only one article each. This study applied Bradford's law and divided journals into three groups by the rank of journals, with each group of journals publishing approximately the same number of articles (Table 4). The top 20 journals represented one-third of all articles. In contrast, 113 and 708 journals contributed to the next two one-third proportions, respectively. The ratio of journal numbers among these three groups is 20:113:708 (1:5.7:35.4), which is quite close to 1:5.7:5.7^2 (32.5).

Table 4. Scattering of articles in journals.

Group	Journals (n=841), n	Articles (n=4473), n	Cumulative, n (%)	Description
Top third	20	1504	1504 (33.6)	Publishing 26-419 articles
Middle third	113	1315	2819 (63.0)	Publishing 8-25 articles
Bottom third	708	1654	4473 (100.0)	Publishing 1-7 articles

Role of Open Access Mega Journals

Open access journals published 35% of the studies, whereas OAMJs published around one-fifth of the studies (Table 1). Half of the top 20 journals are open access journals, with 4 journals considered to be OAMJs (Table 5). The four OAMJs (ie, PLOS ONE, Medicine, Scientific Reports, and BMJ Open) ranked first, second, seventh, and ninth, respectively, in terms of the number of publications. Table 6 shows the distribution of study types per article title across journal types. The proportions of study types were similar between open access journals and non-open access journals (*P*=.266). In contrast, OAMJs had a different distribution of study types than non-OAMJs (*P*<.001). They tended to publish articles with ≥2 medical conditions in the title.

Table 5. Top 20 journals ranked by published articles between 2000 and 2017.

Journal name	IF[a]	Rank in JCR[b] in 2017	Articles, n (%)	OAJ[c]	OAMJ[d]
PLOS ONE	2.766	Multidisciplinary sciences (15/64)	419 (9.4)	Y	Y
Medicine	2.028	Medicine, general & internal (56/154)	385 (8.6)	Y	Y[e]
International Journal of Cardiology	4.034	Cardiac & cardiovascular systems (41/128)	72 (1.6)	N	N
Journal of the Formosan Medical Association	2.452	Medicine, general & internal (42/154)	52 (1.2)	Y	N
Oncotarget	N/A[f]	N/A	51 (1.1)	Y[g]	N
Journal of Affective Disorders	3.786	Clinical neurology (46/197), psychiatry (37/142), psychiatry[h] (27/142)	51 (1.1)	N	N
Scientific Reports	4.122	Multidisciplinary sciences (12/64)	49 (1.1)	Y	Y
Pharmacoepidemiology and Drug Safety	2.314	Public, environmental & occupational health (66/180), pharmacology & pharmacy (145/261)	47 (1.1)	N	N
BMJ Open	2.413	Medicine, general & internal (43/154)	47 (1.1)	Y	Y
Journal of the Chinese Medical Association	1.660	Medicine, general & internal (72/154)	39 (0.9)	Y	N
Journal of Ethnopharmacology	3.115	Plant sciences (38/222); chemistry, medicinal (20/59); integrative & complementary medicine (4/27); pharmacology & pharmacy (87/261)	36 (0.8)	N	N
BMC Health Services Research	1.843	Health care sciences & services (53/94)	34 (0.8)	Y	N
European Journal of Internal Medicine	3.282	Medicine, general & internal (27/154)	30 (0.7)	N	N
Research in Developmental Disabilities	1.820	Education, special[h] (8/40), rehabilitation[h] (19/69)	30 (0.7)	N	N
Health Policy	2.293	Health care sciences & services (40/94), health policy & services[h] (22/79)	29 (0.6)	N	N
Osteoporosis International	3.856	Endocrinology & metabolism (40/143)	28 (0.6)	N	N
Evidence-based Complementary and Alternative Medicine	2.064	Integrative & complementary medicine (10/27)	27 (0.6)	Y	N
QJM	3.204	Medicine, general & internal (30/154)	27 (0.6)	N	N
Journal of Clinical Psychiatry	4.247	Psychiatry (26/142), psychiatry[h] (19/142), psychology, clinical[h] (11/127)	26 (0.6)	N	N
International Journal of Environmental Research and Public Health	2.145	Environmental sciences (116/241); public, environmental & occupational health (73/180); public, environmental & occupational health[h] (44/156)	25 (0.6)	Y	N

[a] IF: impact factor.

[b] JCR: Journal Citation Reports.

[c] OAJ: open access journal.

[d] OAMJ: open access mega journal.

[e] Converted to an OAMJ in 2014.

[f] N/A: not available.

[g] Not listed in the Directory of Open Access Journals.

[h] Social Sciences Edition.

Table 6. Distribution of study types per article title across journal types.

Study type	Open access journal, n (%)		Open access mega journal, n (%)	
	Yes (n=1566)	No (n=2907)	Yes (n=906)	No (n=276)
With ≥1 intervention	552 (35.2)	994 (34.2)	308 (34.0)	1238 (34.7)
With ≥2 conditions	537 (34.3)	1038 (35.7)	400 (44.2)	1175 (32.9)
With only 1 condition	353 (22.5)	684 (23.5)	159 (17.5)	878 (24.6)
Others	124 (7.9)	191 (6.6)	39 (4.3)	276 (7.7)

Prolific Authors and Research Groups

The visualization in Figure 4 presents the networks of co-authorships during 2000 to 2005, 2006 to 2011, and 2012 to 2017. To avoid cluttering the figure, only author pairs who had collaborations in more than 2, 5, and 20 articles in the 3 time periods, respectively, were depicted. From 2000 to 2005, the social network analysis identified three main research groups. Herng-Ching Lin from Taipei Medical University was the most productive author and published at least 10 articles. Between 2006 and 2011, Herng-Ching Lin still produced the highest volume of publications (≥100). However, several other research groups emerged during this period. From 2012 to 2017, a total of 7 productive authors were identified, and each published at least 100 articles. Two large research groups can be readily observed in the network graph, including authors mainly from China Medical University Hospital (Cheng-Li Lin, Chia-Hung Kao, Fung-Chang Sung, et al) and Taipei Veterans General Hospital (Tseng-Ji Chen, Chia-Jen Liu, et al).

When a prolific author was defined as one who had at least 100 articles published between 2000 and 2017, a total of 8 authors were qualified as prolific. Table 7 shows that study types were significantly different between studies authored by at least one of the prolific authors and those that were not ($P<.001$). The most common type of studies contributed by prolific authors were studies with two or more medical conditions in the article title, whereas studies by nonprolific authors were more likely to mention at least one intervention in the article title.

Figure 4. Co-authorship networks during (A) 2000-2005, (B) 2006-2011, and (C) 2012-2017.

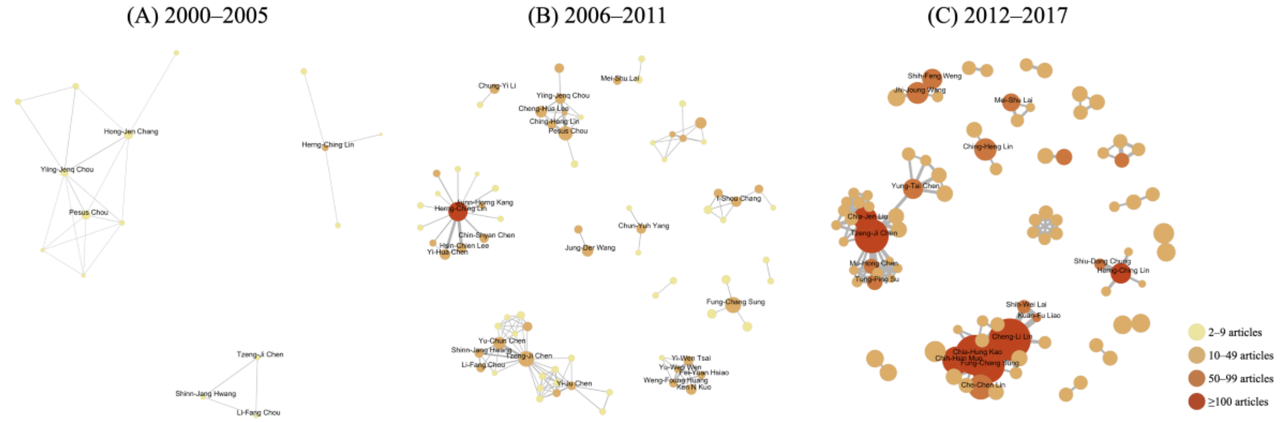


Table 7. Distribution of study types per article title between studies authored by prolific authors and those authored by others.

Study type	Prolific author (≥100 articles), n (%)	
	Yes (n=1399)	No (n=3074)
With ≥1 intervention	385 (27.5)	1161 (37.8)
With ≥2 conditions	707 (50.5)	868 (28.2)
With only 1 condition	250 (17.9)	787 (25.6)
Others	57 (4.1)	258 (8.4)

Discussion

Principal Findings

This study yielded several interesting findings. First, a rapid growth in publications was observed from 2009 to 2015, just as it was between 2000 and 2009. However, the growth dramatically ceased after 2015. Second, certain medical conditions, such as diabetes, stroke, and dementia, and certain interventions such as statin therapy, metformin, and Chinese herbal medicine, received more attention from researchers using NHI claims data as the study material. Third, almost all the studies were published in JCR-indexed journals, most ranking as Q1 or Q2 in their corresponding JCR categories. OAMJs appeared to provide fertile soil for the rapid growth of research based on NHI claims data, particularly studies with ≥2 medical conditions in the article title. Fourth, while the top 8 most prolific authors contributed nearly one-third of all studies, they published more studies with ≥2 medical conditions in the article title than nonprolific authors. These studies mainly investigated the association between two medical conditions and might be easier to conduct than studies examining the effect of an intervention on a medical condition.

Publication Volume

As described by Chen et al [8], the growth of literature using NHI claims data generally followed the proposed model of scientific growth [21] before 2015. The sudden halt of the growth trend after 2015 is quite unexpected. Although the underlying reasons are not fully understood, it is speculated to be related to the 2012 lawsuit against the secondary use of health insurance data [12,13]. Later, in reaction to the legal controversy, more precautionary measures for accessing NHI claims data were enforced [13]. Consequently, the National Health Research Institutes stopped the acceptance of new applications for procuring data from the NHIRD from December 2015 onwards. Moreover, permission to use a local copy of the NHIRD typically expires after 3 years. Therefore, after December 2018, researchers were allowed to access NHI claims data, which is now a part of the Health and Welfare Database, only within the Data Science Center of the Ministry of Health and Welfare or via a virtual private network from local branch offices of the Data Science Center across the country [22]. These measures definitely increased the barriers to conducting research using the NHI claims data. The question of whether the research output will decline from the present level awaits further observation.

Research Topics

Diabetes, stroke, and dementia represented the most commonly investigated medical conditions. All these conditions are highly prevalent diseases that naturally attract more attention from researchers. Furthermore, their high prevalence enabled researchers to study these diseases using merely the Longitudinal Health Insurance Database, which is the 1 million–person subset of the NHIRD that entails a lower cost than the whole dataset of the NHIRD. In particular, because the diagnostic codes for diabetes and stroke have been validated within NHI claims data [23-25], researchers might be more confident performing research on these diseases. This again emphasizes the importance of case validation in secondary data analysis [26].

As for interventions, statins and metformin are commonly prescribed to patients with stroke and diabetes, respectively. Naturally, they were among the most frequently investigated interventions. In addition, the pleiotropic effects of statins and metformin might also intrigue researchers to test their effects on other diseases using large health care databases like the NHIRD [27,28]. Last but not least, the NHI program reimburses Chinese herbal medicine; therefore, NHI claims data provide researchers a unique opportunity to study the effectiveness of Chinese herbal medicine [29].

Implications

Writing for publication is essential for academics. Currently, not only are academics evaluated against how well they publish but universities are also ranked according to their academic publication rates. The long-existing “publish or perish” culture of academia has now prevailed in Taiwan’s hospitals. Taiwan’s hospital accreditation system, in addition to assessing the quality of health care, also aims to determine the teaching status of a hospital [30]. Therefore, hospitals are putting more pressure on their staff to publish in JCR-indexed journals in order to meet the accreditation standards. The pressure to publish is reflected in the increasing trend of first authors coming from hospitals (Table 1).

In addition to these internal factors, the external environment is just suitable for catalyzing the growth of publications. The increasing availability of open access journals, in particular OAMJs, provides unprecedented capacity to accommodate a large volume of publications. Furthermore, several OAMJs (eg, PLOS ONE, Medicine, Scientific Reports, and BMJ Open) have decent impact factors and above-average JCR ranking (Q1 or Q2). All these factors have driven researchers to utilize secondary data analysis to augment their research output. Although the current system might have misdirected some hospital practitioners to “shallow research,” it has also encouraged positive involvement of practitioners in academic research.

Based on the text mining analysis, prolific authors tended to produce articles with ≥2 medical conditions in the title, while such articles were more likely to be published in OAMJs. From the pragmatic point of view, it is easier to investigate the association between two medical conditions than to study the effect of an intervention on a medical condition, in particular when the intervention, such as a medication, is time-dependent. Testing multiple hypotheses at the same time definitely increases the likelihood of finding an association [31]. Therefore, conducting studies investigating the association between two medical conditions within a large database appears to be a shortcut to increase research output. A typical approach is to examine whether condition A increases the risk of condition B. Some of these studies are criticized as “templated and non-hypothesis driven” and have raised concerns and disputes regarding the misuse of data analysis [9,32,33]. In addition, due to Berkson’s bias, such association studies may generate significant but spurious associations caused by inappropriate conditional factors such as hospitalization [34]. All these criticisms may give an impression that findings from studies using NHI claims data are useless to either clinical practitioners or policy makers.

Future Directions

Despite the negative impression, the following strategies were proposed to increase the impact of research based on NHI claims data. First, the percentage of first authors who are not Taiwanese citizens was very low (Table 1). As compared to studies using the United Kingdom’s electronic health database [35], international collaboration is relatively uncommon in studies using NHI claims data. In addition to seeking collaboration with foreign partners, the integration of databases from multiple countries can provide opportunities to compare disease prevalence and treatment effects across countries [36], hopefully producing more generalizable knowledge. Second, apart from formulating a hypothesis before conducting a study, researchers should focus on generating research questions with a real impact on clinical decision making rather than producing articles acceptable by journals with an impact factor. In this respect, studies using the United Kingdom's electronic health database have set good examples of providing real-world evidence to inform clinical practice [37]. Future research should be directed towards clinically relevant and actionable study outcomes. Third, NHI claims data, like other administrative claims data, have been questioned about their data validity because claims data typically lack clinical information. Record linkage between NHI claims data and various clinical registries may be a promising approach to complement the advantages of each data source. The combination of detailed clinical information from registry databases and long-term outcome data from claims databases offers opportunities to enhance the validity of outcomes research [38]. Finally, as for the litigation from human rights groups against the use of NHI claims data, researchers should take this as an opportunity to meet the need for more dialogue and proactive participation from such groups.

Limitations

This study has the following limitations. First, this study included only articles written in English and indexed in the PubMed database to make the results comparable with the study by Chen et al [8]. This undoubtedly limited the scope of this work to health-related publications. However, because the purpose of NHI claims data is to track payments for health care utilization, it is believed that most of these articles, if not all, were published in PubMed-indexed journals. Second, even though journal rankings change year by year, this study used 2017 JCR Journal Impact Factors and journal categories for journal ranking to simplify the analysis. Third, the analysis of article titles was based on automated natural language processing algorithms provided by MetaMap. Although MetaMap can effectively extract medical concepts from biomedical texts, the semantic relationships between the identified medical concepts are not readily discernible [39]. For example, it was difficult to determine whether a study truly investigated the association between two medical conditions simply based on the co-occurrence of the two conditions in the article title.

Conclusions

As Taiwan has recently become an aged society and is expected to become a super-aged society by 2025 [40,41], health care expenditure will inevitably increase as the size of the elderly population grows. Therefore, knowledge of the burden of various diseases as well as the cost effectiveness of different diagnostic and treatment strategies is of paramount importance. Although nationwide disease registration and population-based surveys can provide valuable information to facilitate medical decisions and policy making, the processes of registration and surveys may themselves entail extra costs. In this regard, analysis of secondary data, such as NHI claims data, may continue to provide an affordable, alternative means of gaining knowledge.

Acknowledgments

We would like to thank Ms. Li-Ying Sung for English language editing. This research was supported in part by the Ministry of Science and Technology (grant number MOST 107-2314-B-705-001) and the Center for Innovative Research on Aging Society from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Flowchart of included articles.

[PDF File (Adobe PDF File), 313 KB - jmir_v22i6e18457_app1.pdf]

Multimedia Appendix 2

Unified Medical Language System (UMLS) semantic types for categorizing medical entities.

[PDF File (Adobe PDF File), 62 KB - jmir_v22i6e18457_app2.pdf]

References

1. Cadarette SM, Wong L. An Introduction to Health Care Administrative Data. Can J Hosp Pharm 2015;68(3):232-237 [FREE Full text] [doi: 10.4212/cjhp.v68i3.1457] [Medline: 26157185]
2. Ferver K, Burton B, Jesilow P. The Use of Claims Data in Healthcare Research. Open Publ Health J 2009 Apr 02;2(1):11-24. [doi: 10.2174/1874944500902010011]
3. Cheng S, Jin H, Yang B, Blank RH. Health Expenditure Growth under Single-Payer Systems: Comparing South Korea and Taiwan. Value Health Reg Issues 2018 May;15:149-154. [doi: 10.1016/j.vhri.2018.03.002] [Medline: 29730247]

4. Hallas J, Wang SV, Gagne JJ, Schneeweiss S, Pratt N, Pottegård A. Hypothesis-free screening of large administrative databases for unsuspected drug-outcome associations. Eur J Epidemiol 2018 Jun;33(6):545-555. [doi: 10.1007/s10654-018-0386-8] [Medline: 29605890]
5. Tricco AC, Pham B, Rawson NSB. Manitoba and Saskatchewan administrative health care utilization databases are used differently to answer epidemiologic research questions. J Clin Epidemiol 2008 Feb;61(2):192-197. [doi: 10.1016/j.jclinepi.2007.03.009] [Medline: 18177793]
6. Hsing AW, Ioannidis JPA. Nationwide Population Science: Lessons From the Taiwan National Health Insurance Research Database. JAMA Intern Med 2015 Sep;175(9):1527-1529. [doi: 10.1001/jamainternmed.2015.3540] [Medline: 26192815]
7. Chen Y, Wu J, Chen T, Wetter T. Reduced access to database. A publicly available database accelerates academic production. BMJ 2011 Feb 01;342:d637. [doi: 10.1136/bmj.d637] [Medline: 21285219]
8. Chen Y, Yeh H, Wu J, Haschler I, Chen T, Wetter T. Taiwan's National Health Insurance Research Database: administrative health care database as study object in bibliometrics. Scientometrics 2011;86(2):365-380. [doi: 10.1007/s11192-010-0289-2]
9. Hampson NB, Weaver LK. Carbon monoxide poisoning and risk for ischemic stroke. Eur J Intern Med 2016 Jun;31:e7. [doi: 10.1016/j.ejim.2016.01.006] [Medline: 26809864]
10. Wakeling S, Willett P, Creaser C, Fry J, Pinfield S, Spezi V. Open-Access Mega-Journals: A Bibliometric Profile. PLoS One 2016;11(11):e0165359 [FREE Full text] [doi: 10.1371/journal.pone.0165359] [Medline: 27861511]
11. Wakeling S, Willett P, Creaser C, Fry J, Pinfield S, Spezi V. Transitioning from a Conventional to a 'Mega' Journal: A Bibliometric Case Study of the Journal Medicine. Publications 2017 Apr 06;5(2):7. [doi: 10.3390/publications5020007]
12. Chang C. Controversy over Information Privacy Arising from the Taiwan National Health Insurance Database Examining the Taiwan Taipei High Administrative Court Judgment No. 102-SU-36 (Tsai v. NHIA). Pace International Law Review 2016;28(1):29-116.
13. Tsai F, Junod V. Medical research using governments' health claims databases: with or without patients' consent? J Public Health (Oxf) 2018 Dec 01;40(4):871-877. [doi: 10.1093/pubmed/fdy034] [Medline: 29506041]
14. National Health Research Institutes. National Health Insurance Research Database URL: https://nhird.nhri.org.tw/en/index.html [accessed 2018-10-04]
15. Hsieh C, Su C, Shao S, Sung S, Lin S, Kao Yang Y, et al. Taiwan's National Health Insurance Research Database: past and future. Clin Epidemiol 2019;11:349-358 [FREE Full text] [doi: 10.2147/CLEP.S196293] [Medline: 31118821]
16. Lai EC, Man KKC, Chaiyakunapruk N, Cheng C, Chien H, Chui CSL, et al. Brief Report: Databases in the Asia-Pacific Region: The Potential for a Distributed Network Approach. Epidemiology 2015 Nov;26(6):815-820. [doi: 10.1097/EDE.0000000000000325] [Medline: 26133022]
17. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
18. Abacha A, Zweigenbaum P. Medical entity recognition: A comparison of semantic and statistical methods. Stroudsburg, PA 18360, United States: Association for Computational Linguistics; 2011 Presented at: Proceedings of BioNLP 2011 Workshop; 24 June, 2011; Portland, Oregon, USA p. 56-64.
19. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002 Jan 05;359(9300):57-61. [doi: 10.1016/S0140-6736(02)07283-5] [Medline: 11809203]
20. Szabo A. Test for Trend With a Multinomial Outcome. The American Statistician 2019;73(4):313-320. [doi: 10.1080/00031305.2017.1407823]
21. Szydłowski M, Krawiec A. Growth cycles of knowledge. Scientometrics 2008 Dec 20;78(1):99-111. [doi: 10.1007/s11192-007-1958-7]
22. Hsieh C, Su C, Shao S, Lin S, BPharm Y, Lai E. The Health and Welfare Database of Taiwan - Opportunities for High Quality Real-World Evidence from Big Data Analyses. Pharmaceutical and medical device regulatory science 2018;49(7):425-436.
23. Cheng C, Kao YY, Lin S, Lee C, Lai ML. Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. Pharmacoepidemiol Drug Saf 2011 Mar;20(3):236-242. [doi: 10.1002/pds.2087] [Medline: 21351304]
24. Hsieh C, Chen C, Li C, Lai M. Validating the diagnosis of acute ischemic stroke in a National Health Insurance claims database. J Formos Med Assoc 2015 Mar;114(3):254-259 [FREE Full text] [doi: 10.1016/j.jfma.2013.09.009] [Medline: 24140108]
25. Lin C, Lai M, Syu C, Chang S, Tseng F. Accuracy of diabetes diagnosis in health insurance claims data in Taiwan. J Formos Med Assoc 2005 Mar;104(3):157-163. [Medline: 15818428]
26. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005 Apr;58(4):323-337. [doi: 10.1016/j.jclinepi.2004.10.012] [Medline: 15862718]
27. Chuang M, Yang Y, Tsai Y, Hsieh M, Lin Y, Lin C, et al. Survival benefit associated with metformin use in inoperable non-small cell lung cancer patients with diabetes: A population-based retrospective cohort study. PLoS One 2018;13(1):e0191129 [FREE Full text] [doi: 10.1371/journal.pone.0191129] [Medline: 29329345]
28. Huang Y, Lee C, Yang S, Fu S, Chen Y, Wang T, et al. Statins Reduce the Risk of Cirrhosis and Its Decompensation in Chronic Hepatitis B Patients: A Nationwide Cohort Study. Am J Gastroenterol 2016 Jul;111(7):976-985. [doi: 10.1038/ajg.2016.179] [Medline: 27166128]

29. Tsai T, Livneh H, Hung T, Lin I, Lu M, Yeh C. Associations between prescribed Chinese herbal medicine and risk of hepatocellular carcinoma in patients with chronic hepatitis B: a nationwide population-based cohort study. BMJ Open 2017 Jan 25;7(1):e014571 [FREE Full text] [doi: 10.1136/bmjopen-2016-014571] [Medline: 28122837]
30. Huang P, Hsu YH, Kai-Yuan T, Hsueh YS. Can European external peer review techniques be introduced and adopted into Taiwan's hospital accreditation system? Int J Qual Health Care 2000 Jun;12(3):251-254. [doi: 10.1093/intqhc/12.3.251] [Medline: 10894199]
31. Austin PC, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. J Clin Epidemiol 2006 Sep;59(9):964-969. [doi: 10.1016/j.jclinepi.2006.01.012] [Medline: 16895820]
32. Kao C. Clinical studies stemming from the Taiwan National Health Insurance Database. Eur J Intern Med 2016 Jun;31:e8. [doi: 10.1016/j.ejim.2016.02.002] [Medline: 26897658]
33. Kao C. Author's Reply: National Health Insurance database in Taiwan. Eur J Intern Med 2016 Jun;31:e11-e12. [doi: 10.1016/j.ejim.2016.03.032] [Medline: 27085390]
34. Wu Y, Lee H. National Health Insurance database in Taiwan: A resource or obstacle for health research? Eur J Intern Med 2016 Jun;31:e9-e10. [doi: 10.1016/j.ejim.2016.03.022] [Medline: 27079475]
35. Chen Y, Wu J, Haschler I, Majeed A, Chen T, Wetter T. Academic impact of a public electronic health database: bibliometric analysis of studies using the general practice research database. PLoS One 2011;6(6):e21404 [FREE Full text] [doi: 10.1371/journal.pone.0021404] [Medline: 21731733]
36. Lai EC, Ryan P, Zhang Y, Schuemie M, Hardy NC, Kamijima Y, et al. Applying a common data model to Asian databases for multinational pharmacoepidemiologic studies: opportunities and challenges. Clin Epidemiol 2018;10:875-885 [FREE Full text] [doi: 10.2147/CLEP.S149961] [Medline: 30100761]
37. Oyinlola JO, Campbell J, Kousoulis AA. Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. BMC Health Serv Res 2016 Jul 26;16:299 [FREE Full text] [doi: 10.1186/s12913-016-1562-8] [Medline: 27456701]
38. Hsieh C, Wu DP, Sung S. Registry-based stroke research in Taiwan: past and future. Epidemiol Health 2018;40:e2018004 [FREE Full text] [doi: 10.4178/epih.e2018004] [Medline: 29421864]
39. Ogallo W, Kanter AS. Using Natural Language Processing and Network Analysis to Develop a Conceptual Framework for Medication Therapy Management Research. AMIA Annu Symp Proc 2016;2016:984-993 [FREE Full text] [Medline: 28269895]
40. Hsu M, Liao P. Financing National Health Insurance: The Challenge of Fast Population Aging. Taiwan Economic Review 2015;43(2):145-182.
41. Lin Y, Huang C. Aging in Taiwan: Building a Society for Active Aging and Aging in Place. Gerontologist 2016 Apr;56(2):176-183. [doi: 10.1093/geront/gnv107] [Medline: 26589450]

Abbreviations

JCR: Journal Citation Reports.
MeSH: medical subject heading.
NHI: National Health Insurance.
NHIRD: National Health Insurance Research Database.
OAJ: open access journal.
OAMJ: open access mega journal.
UMLS: Unified Medical Language System.

Edited by G Eysenbach; submitted 27.02.20; peer-reviewed by PJ Lee, M Torii; comments to author 06.04.20; revised version received 12.04.20; accepted 16.04.20; published 16.06.20.

Please cite as:
Sung SF, Hsieh CY, Hu YH
Two Decades of Research Using Taiwan’s National Health Insurance Claims Data: Bibliometric and Text Mining Analysis on PubMed
J Med Internet Res 2020;22(6):e18457
URL: http://www.jmir.org/2020/6/e18457/
doi:10.2196/18457
PMID:32543443

Review

Current Evidence for Continuous Vital Signs Monitoring by Wearable Wireless Devices in Hospitalized Adults: Systematic Review

Jobbe P L Leenen[1], MSc, RN; Crista Leerentveld[2], MSc, RN; Joris D van Dijk[3], MSc, PhD; Henderik L van Westreenen[1], MD, PhD; Lisette Schoonhoven[4,5], PhD; Gijsbert A Patijn[1], MD, PhD

[1]Department of Surgery, Isala, Zwolle, Netherlands
[2]Department of Intensive Care, Isala, Zwolle, Netherlands
[3]Isala Academy, Isala, Zwolle, Netherlands
[4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands
[5]School of Health Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Jobbe P L Leenen, MSc, RN
Department of Surgery
Isala
Dr. van Heesweg 2
Zwolle, 8025 AB
Netherlands
Phone: 31 384245654
Email: j.p.l.leenen@isala.nl

Abstract

Background: Continuous monitoring of vital signs by using wearable wireless devices may allow for timely detection of clinical deterioration in patients in general wards in comparison to detection by standard intermittent vital signs measurements. A large number of studies on many different wearable devices have been reported in recent years, but a systematic review is not yet available to date.

Objective: The aim of this study was to provide a systematic review for health care professionals regarding the current evidence about the validation, feasibility, clinical outcomes, and costs of wearable wireless devices for continuous monitoring of vital signs.

Methods: A systematic and comprehensive search was performed using PubMed/MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials from January 2009 to September 2019 for studies that evaluated wearable wireless devices for continuous monitoring of vital signs in adults. Outcomes were structured by validation, feasibility, clinical outcomes, and costs. Risk of bias was determined by using the Mixed Methods Appraisal Tool, quality assessment of diagnostic accuracy studies 2nd edition, or quality of health economic studies tool.

Results: In this review, 27 studies evaluating 13 different wearable wireless devices were included. These studies predominantly evaluated the validation or the feasibility outcomes of these devices. Only a few studies reported the clinical outcomes with these devices and they did not report a significantly better clinical outcome than the standard tools used for measuring vital signs. Cost outcomes were not reported in any study. The quality of the included studies was predominantly rated as low or moderate.

Conclusions: Wearable wireless continuous monitoring devices are mostly still in the clinical validation and feasibility testing phases. To date, there are no high quality large well-controlled studies of wearable wireless devices available that show a significant clinical benefit or cost-effectiveness. Such studies are needed to help health care professionals and administrators in their decision making regarding implementation of these devices on a large scale in clinical practice or in-home monitoring.

(J Med Internet Res 2020;22(6):e18636) doi:10.2196/18636

KEYWORDS

continuous monitoring; patient monitoring; vital signs; clinical deterioration; early deterioration; wearable wireless device; systematic review; monitoring

Introduction

Continuous monitoring of vital signs of inpatients is a common practice in intensive care, medium care, operation theatre, and recovery ward settings [1]. The goal of continuous vital signs monitoring in these settings is early detection of the clinical deterioration, thereby allowing timely intervention [2,3]. However, once patients are discharged to the general ward, vital signs are only monitored intermittently, often just once or twice daily. Early warning scores have been implemented to guide clinical interpretation, but this value is limited by the intermittent nature of the measurements [4-6]. Serious unexpected adverse events do occur regularly in general wards, especially in high-risk postsurgical or elderly frail patients [7-13]. This incidence of adverse events is expected to increase owing to the aging population, increasing complexity of in-hospital care, increasing pressure to limit health care costs, and increasing shortage of nursing staff. These adverse events may be prevented or mitigated if continuous monitoring of vital signs would be available to facilitate early detection of the deteriorating trends in vital signs, thereby allowing timely interventions [14-16]. An important advantage of continuous monitoring may be the insight in the trends, which can be much more informative and predictive than single deviating values [17-19].

Recent studies have shown that continuous monitoring in combination with automated alerts in case of deterioration improves patient outcomes [17,20-23]. However, for continuous monitoring to be applicable in general wards, it should not lead to decreased mobility of the patient. Therefore, continuous monitoring devices should preferably be portable, wireless, and wearable on an easily accessible body part [18,24]. Such wearable devices also have the potential to be used for continuous monitoring of the vital signs of the patients at home or in rehabilitation centers, thereby possibly leading to reduced length of hospital stay and preventing unplanned readmissions [25].

The technology of wearable wireless sensors for vital signs monitoring is advancing rapidly [26]. Many manufacturers are now developing wearable sensors with different capabilities and different underlying technical specifications and algorithms [27]. The reliability and the accuracy of these devices have often only been demonstrated in healthy volunteers instead of in patients with deviating values [17]. In addition, the scientific evidence regarding the feasibility, effectiveness, and costs of these wearable sensors in clinical practice is still very limited [17,28,29]. Previous reviews on continuous monitoring of vital signs did not focus on wearable wireless devices but rather on conventional nonambulant monitoring [14]. The aim of this study was to systematically review the current evidence on wearable wireless devices for continuous vital signs monitoring by providing a thorough overview of the currently available studies.

Methods

Design

We conducted a systematic review of the literature by following the guidelines as outlined in the Cochrane Handbook for Systematic Reviews of Interventions version 6.0 and reported according the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [30,31].

Eligibility Criteria

Studies were considered eligible for inclusion when they met the following criteria: consisted of participants older than 18 years; evaluated a continuous monitoring device that measured vital signs such as heart rate (HR), respiratory rate (RR), blood pressure (BP), temperature, and blood oxygen saturation (SpO_2) [16]; used a device that measured ≥2 vital signs; used a device that was wireless and wearable; and published after 2009. This timeframe was chosen to prevent the inclusion of papers on outdated technology. Studies were excluded when the device was not wearable by the patient and the device had no formal approval as a medical device through the Conformité Européenne (CE) mark or Food and Drug Administration (FDA) clearance or both. Furthermore, conference abstracts, review articles, letters, editorials, articles without full texts, and non-English or non-Dutch articles were excluded.

The outcomes of interest were as follows: validation (eg, sensitivity, specificity, limits of agreement [LoA]), feasibility (eg, acceptability, user experiences, system fidelity), clinical outcomes (eg, mortality, length of stay, fail-to-rescue [FTR], intensive care unit [ICU] admission), and costs (eg, cost-minimization, cost-benefit, cost-effectiveness, or cost-utility outcomes) [25,32-35].

For validation studies, the prespecified clinically relevant mean difference and LoA were 10±10 beats per minute for HR, 3±3 breaths per minute for RR, 0.5°C±1.0°C for temperature, 10±20 mmHg for systolic BP, and 3%±5% for SpO_2. The guidelines for the acceptable mean differences and LoA for continuous monitoring of vital signs are unfortunately lacking.

Search Strategy

A systematic literature search of PubMed/MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials was performed with the last search run on September 6, 2019. In addition, the references of the retrieved studies were manually screened to obtain additional relevant studies. The following keywords were used: vital signs, clinical deterioration, and wireless continuous monitoring. Keywords on outcomes were based on terms about validation, feasibility, clinical outcomes, and cost outcomes. The full search strategy is available in Multimedia Appendix 1. The search string was audited by a clinical librarian and adapted for the individual databases and interfaces as needed. The information about the specifications of the wearable devices was obtained from the manuals and fact sheets of the manufacturers.

Study Selection

All identified references were checked for duplicates and consolidated in the reference manager software (Mendeley 1.19.5). Titles and abstracts of references were independently screened by 2 researchers against the inclusion and exclusion criteria. Full-text articles of references that matched the inclusion criteria were read independently to determine eligibility. Disagreements were resolved by discussion between the 2 review authors; if no agreement could be reached, the third author was consulted.

Data Collection Process

A data extraction sheet was developed based on the Cochrane Consumers and Communication review group's data extraction template and was pilot tested using 5 randomly selected included studies and refined accordingly [31]. One review author extracted the data from the included studies and the second author checked the extracted data. Disagreements were resolved by discussion between the 2 review authors; if no agreement could be reached, the third author was consulted.

Data Extraction and Synthesis

The following data were extracted for each study: (1) first author, country, year of publishing, aim, design, setting, patient population, sample size, and conflicts of interest; (2) manufacturer and name of the device and type of vital signs measured by the device; and (3) outcomes of the studies divided in previously defined categories: validation, feasibility, clinical, and cost outcomes. The study outcomes were presented for each device.

Risk of Bias of Individual Studies

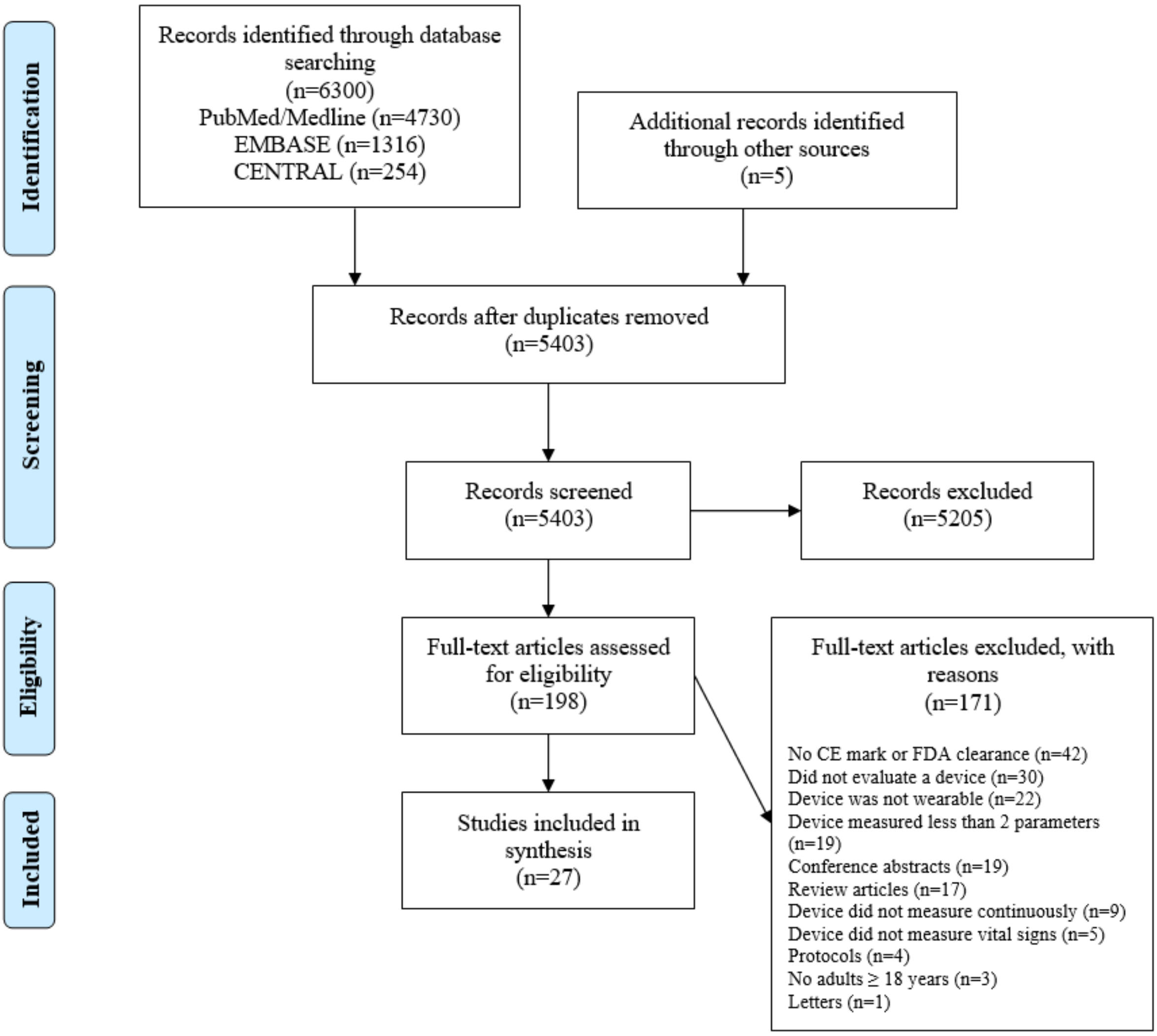
For assessing the risk of bias of individual studies, 2 authors independently appraised each study critically. Disagreements in the quality assessment between the authors were solved by discussion until consensus was reached. Owing to the large diversity of the included study designs, 3 different instruments were used. The 2018 version of the Mixed Methods Appraisal Tool (MMAT) was utilized for 5 study designs: qualitative, quantitative randomized controlled, quantitative nonrandomized, quantitative descriptive, and mixed methods [36]. Each category contained 5 criteria with the score range from 0 to 5 of the criteria met. For mixed methods studies, scores were calculated as the lowest score from among the 3 relevant designs (quantitative, qualitative, and mixed methods). A score of 0 to 2 was considered as low, a score of 3 and 4 was considered as moderate, and a score of 5 was considered as high. For diagnostic accuracy study designs, the quality assessment of diagnostic accuracy studies 2nd edition (QUADAS-2) was utilized to assess the risk of bias [37]. QUADAS‐2 consists of 4 domains: patient selection, index test, reference standard, and flow and timing. All domains were assessed for the potential for risk of bias and the first 3 domains, that is, patient selection, index test, and reference standard were assessed for concerns regarding applicability. For economic evaluation studies, the quality of health economic studies (QHES) tool was utilized to assess the quality [38]. The QHES instrument is a validated method for assessing the quality of health economic analyses. It consists of 16 items, each with specific weight values ranging from 1 to 9. Each score is multiplied by the weight to produce a total score, with a maximum score of 100.

Results

Study Selection

We identified 5403 potentially relevant studies in our literature search after duplicate removal, of which 5 studies were accessed from the reference list of the potentially relevant studies. Screening of titles and abstracts resulted in 198 studies, which were read full text. Eventually, 27 studies that met the eligibility criteria were included [39-65]. A PRISMA flowchart of the search is presented in Figure 1.

Figure 1. PRISMA flowchart. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; CENTRAL: Cochrane Central Register of Controlled Trials; CE: Conformité Européenne; FDA: Food and Drug Administration.



Study Characteristics

In this study, 13 different devices of 10 manufacturers were studied in 2717 subjects (median 43, range 6-736). Subjects were healthy patients, trauma patients, surgical patients, or neurological/neurosurgical patients (Table 1). The 13 devices were as follows: ViSi Mobile, SensiumVitals, HealthPatch MD, VitalPatch, Wireless Vital Signs Monitor (WVSM) device, MiniMedic, Zephyr BioPatch, Biosensor, IntelliVue Cableless Measurement Solution, Wavelet Wristband, Proteus patch, Alarm Management System, and EQ02 Lifemonitor (Table 2).

Of the 27 included studies, 15 were from the United States and the remaining were from the United Kingdom (N=6), the Netherlands (N=2), Canada (N=1), China (N=1), Australia (N=1), and Austria (N=1). Among these, 13 were validation studies, 6 were cohort studies, 2 were case-control studies, 3 were mixed methods studies, 1 was a qualitative study, and 2 were pilot randomized controlled trials. The reported outcomes were validation (N=15), feasibility (N=15), and clinical outcomes (N=6; Table 3). Seventeen studies declared that they had no conflicts of interest. In 6 studies, one or more authors were employees of the manufacturing company of the studied device. The remaining 4 studies did not declare any possible conflicts of interest (Table 1).

Devices

ViSi Mobile

Five studies (N=1308) have been published about the ViSi Mobile (Sotera Wireless; Table 1) [39,40,51,59,60]. This device is worn on the wrist, upper arm, or chest, and it measures HR, RR, BP, SpO_2, and skin temperature (Table 2) [66].

Validation outcomes: This device was validated in 1 study, which reported an acceptable mean difference but wide LoA between the device and manual nurse measurements for HR, RR, and BP (Table 4) [59]. SpO_2 had an acceptable mean difference and LoA.

Of the 27 included studies, 15 were from the United States and the remaining were from the United Kingdom (N=6), the Netherlands (N=2), Canada (N=1), China (N=1), Australia (N=1), and Austria (N=1). Among these, 13 were validation studies, 6 were cohort studies, 2 were case-control studies, 3

were mixed methods studies, 1 was a qualitative study, and 2 were pilot randomized controlled trials. The reported outcomes were validation (N=15), feasibility (N=15), and clinical outcomes (N=6; Table 3). Seventeen studies declared that they had no conflicts of interest. In 6 studies, one or more authors were employees of the manufacturing company of the studied device. The remaining 4 studies did not declare any possible conflicts of interest (Table 1).

Feasibility outcomes: Patients reported the wristband as big or heavy. Four studies reported the perceptions of the health care professionals [39,51,59,60]. Nurses mentioned that this device had a short battery life and poor connection but it reported better insight into the vital signs [59]. Both nurses and physicians felt confident about their ability to identify patients at risk of deterioration but were concerned about the accuracy of the device [39,59]. Besides, physicians were positive about the potential of continuous monitoring, as this device provided reassurance to patients and supported interdisciplinary communication between nurses and physicians [39]. Another study stated that 67% of the nurses were positive about the deployment of continuous monitoring in the ward [51]. All nurses were positive that the monitor provided valuable patient data that increased patient safety [60]. However, they had certain reservations, including the potential decrease in the bedside nurse-patient contact, increase in inappropriate rapid response team (RRT) calls, and possible discomfort for patients wearing the device [39]. Two studies reported system fidelity. The system generated 2 to 10 alarms per patient in a day [40,60], of which one study [60] reported that 92% of the nurses indicated that the number of alarms were appropriate. One study showed that 70% of the artefacts, defined as the noncollected parameters, were caused by connection failure and 74% lasted less than 5 minutes [59].

Clinical outcomes: RRT calls, FTR, unexpected deaths, and ICU transfers were not significantly reduced by continuous monitoring [40,51]. The complication rate was higher in the intermittent monitoring group than in the continuous monitoring group [51]. One study described only 4 alert-initiated interventions in 236 patients [60]. The quality of these studies ranged from low to moderate, as assessed by the MMAT tool, thereby indicating that these studies are subject to bias (Figure 2).

Cost outcomes: None of the studies reported this type of outcome.

SensiumVitals

Five studies (N=371) have been published about the SensiumVitals (Sensium Healthcare; Table 1) [56,61-64]. This is a patch device attached to the chest for continuous monitoring of the HR, RR, and axillary temperature (Table 2) [67].

Validation outcomes: This device was validated in 3 studies. Two studies included surgical patients [63,64] and 1 included healthy volunteers [61]. The results were conflicting. The mean difference between the device and reference standard was acceptable for HR and RR (Table 4). For HR, LoA was acceptable in 2 studies and outside acceptable limits for 1 study. For RR, LoA was wide for all studies. One study [64] reported temperatures outside acceptable ranges. Furthermore, RR was frequently rejected by the algorithm owing to the inaccuracy of the measurement [61,63]. However, the results may be biased owing to the high risk of bias at the reference standards and patient selection (Figure 2). In addition, 2 of the 3 studies [61,63] were authored by the employees of the SensiumVitals manufacturing company and one study was also funded by the manufacturer [61].

Feasibility outcomes: Two studies described the feasibility of this device. One qualitative study showed the patient perceptions [56]. Six themes emerged from the interviews: (1) patients emphasized the importance of nursing contact, (2) patients indicated that they hoped to be disturbed less for night-time observations with the new monitoring system, (3) patients reported high comfort, (4) patients experienced a high sense of security, (5) patients expressed that monitoring could be a solution for the busy nursing staff, and (6) patients expressed reservations about the reliability of the technology such as the data security and system failure. The second study reported that patients were comfortable with the patch and that it enhanced the feeling of safety although 16.4% discontinued the intervention owing to the discomfort before the end of the study [62].

Clinical outcomes: Only one study reported the clinical outcome. In that study, no statistically significant better clinical outcomes for the patch group were seen, possibly owing to the sample size [62]. Notably, the authors reported that an unacceptable high number of alerts were sent to the nurses before adjusting the alarm thresholds. Since the quality of these studies was rated from low to high by the MMAT tool, possible bias is introduced (Figure 2).

Cost outcomes: None of the studies reported this type of outcome.

VitalPatch and HealthPatch MD

Five studies (N=133) have been published on the VitalPatch and its previous version HealthPatch MD, which is not available anymore (VitalConnect; Table 1) [41-43,65]. Of them, one mixed methods study compared the HealthPatch with the ViSi Mobile [59]. This patch device is applied to the chest and measures HR, RR, and ST (Table 2) [68].

Validation outcomes: This device was validated in 4 studies. For HR, the mean difference was acceptable for all studies and LoA was acceptable for 2 studies (Table 4). The mean difference for RR was acceptable; however, all studies reported LoA outside of the preset acceptable range. One study reported a mean absolute error of less than 3 for HR and 1 for RR [65]. All studies were subject to potential bias at patient selection and the reference standard (Table 4).

Feasibility outcomes: The acceptability of this device was reported as high by the majority of the nurses [41]. However, the exact numbers were not reported. Besides, the health care professionals recommended that it was necessary to gain experience with use of the device in clinical practice [41]. Patients reported that the HealthPatch did not restrict them in daily activities. The fidelity of the system was reported in 2 studies, of which one study reported a loss of data of 6% [42].

They compared several thresholds; 63% of the measurements were performed without data loss greater than 2 minutes. In addition, another study reported that more than 50% of all the artefacts lasted for less than 1 minute, and 43% of them lasted for less than 5 minutes [59]. The reasons for these artefacts were wireless signal connection problems or losing skin contact.

None of the studies reported the clinical and cost outcomes.

WVSM Device

Two studies (N=305) evaluated the WVSM device (Athena GTX) in trauma patients (Table 1) [45,69]. This device measures the HR, BP, RR, and SpO_2 continuously and is worn on the chest, upper arm, and fingertips (Table 2) [70].

Feasibility outcomes: One study reported the feasibility outcomes [45]. This study was a posthoc analysis of the previous study of Liu et al [69]. They found at least 75% adequate data for BP, HR, and RR for predicting life-saving interventions (LSIs) [45]. However, the results were subject to bias because of a high risk of bias in the following categories: patient selection and flow and timing (Figure 2).

Clinical outcomes: One study reported the clinical outcomes and showed that the data of this device were accurate in comparison with that shown in a conventional monitor for the determination of LSIs, without periodic loss of signals or other errors [69]. The authors learned during the study that new medical devices to be used for prehospital studies require integration into the local information technology infrastructure. The quality of this study was rated as high (Figure 2).

None of the studies reported the validation and cost outcomes.

MiniMedic

Two studies (N=155) evaluated the MiniMedic (Athena GTX) in trauma patients (Table 1) [49,50]. This device measures the HR, SpO_2, and ST both at the fingertip and in the forehead (Table 2). In addition, a Murphy factor, an injury acuity algorithm that generates a score, can be calculated for triaged patients in need of LSIs [71].

Validation outcomes: One study compared the pulse-wave transit time, a derivate of BP, reported in the device with the BP reported in the conventional monitor and found correlations between them (R^2=0.036, P<.001; Table 3) [50]. Temperature measurements were significantly different between the device and the reference standard and between the fingertip and the forehead sensor of the device. For HR, a mean difference of 3 beats per minute was found between the device and the reference standard (P<.001). For SpO_2, the median difference between the conventional monitor and the fingertip sensor was 0% and that between the conventional monitor and the forehead sensor was 7% (P<.001). However, this study had a high possibility of bias at patient selection (Figure 2). The second study demonstrated that the MiniMedic was capable of computing a single numeric value, the Murphy factor, to summarize the overall patient status and to identify prehospital trauma patients who need LSIs [49].

None of these studies reported the feasibility, clinical, and cost outcomes.

Zephyr BioPatch

Three studies (N=85) have been published about the Zephyr BioPatch (Medtronic Annapolis; Table 1) [46-48]. This is a patch or a patch fixed by a harness on the chest and it measures the HR, RR, and the estimated core temperature (Table 2) [72].

Validation outcomes: Two studies reported the validation outcomes (Table 3). One study was conducted in healthy volunteers during graded exercise and in a hot environment and one was conducted in full-term pregnant women [47,48]. For HR, both studies reported acceptable mean differences but nonacceptable LoA. For RR, one study [47] reported acceptable mean differences but nonacceptable LoA for RR but the other study [48] that also reported acceptable mean differences but nonacceptable LoA for RR was subjected to a high risk of bias at patient selection (Figure 2). Therefore, Boatin et al [47] are the only researchers who have reported acceptable mean differences but nonacceptable LoA for RR.

Feasibility outcomes: Considering the feasibility outcomes, the participants found the patch comfortable (78%), likeable (81%), and useful (97%). Among nurses, 80% of the nurses found the monitor easy to use and 84% would recommend it to patients [47]. Another study reported a retention rate of 88.6% at the end of the 24-hour monitoring period after exclusion of 2 patients with poor electrocardiogram (ECG) signals [46]. Furthermore, the authors interviewed patients and nurses about any challenges wearing the sensors. Both groups did not report any challenges. The quality of the studies was rated as moderate and high (Figure 2).

None of the included studies reported the clinical and cost outcomes.

Biosensor

One study (N=17) reported about the Philips Biosensor, which is a rebrand of the VitalConnect's HealthPatch (Table 1) [57]. This device is able to measure HR, RR, and ST (Table 2) [73].

Validation outcomes: This study only compared the RR of the device with a reference standard. This resulted in acceptable limits of mean difference of 3.5±5.2 breaths per minute and a statistically significant correlation of Spearman's ρ of 0.86. However, results may be biased due to the high risk of bias regarding patient selection and flow and timing (Figure 2). In addition, 2 authors were employees of Philips and the study was funded by the manufacturer.

This study did not report the feasibility, clinical, and cost outcomes.

Wavelet Wristband

One study (N=35) reported about the Wavelet Wristband (Wavelet Health), a watch that monitors HR and RR (Table 1 and Table 2) [52,74].

Validation outcomes: For HR, acceptable mean differences and LoA were found (Table 4). For RR, the LoA was outside of the acceptable limits. However, all aspects of risk of bias were either unclear or high and applicability was low (Figure 2). Besides, 4 authors were former or current employees of the manufacturing company.

This study did not report the feasibility, clinical, and cost outcomes.

Proteus Patch

We found 1 study (N=13) that reported about the Proteus patch (Proteus Digital Health; Table 1) [53]. This device monitors HR, RR, and ST (Table 2) [75].

Feasibility outcomes: In the feasibility study, the patch was able to monitor for over 5 days at home. However, data of 2 patients was insufficient for performing the analysis and were excluded. The quality of the study was rated as low (Figure 2).

This study did not report the validation, clinical, or cost outcomes.

IntelliVue Cableless Measurement Solution

We found about the IntelliVue Cableless Measurement Solution (Philips) in 1 study on clinical patients (N=226; Table 1) [54]. This is a device for monitoring the HR, RR, BP, and SpO_2 (Table 2) [76].

Feasibility outcomes: There was an overall good acceptance by patients and health care professionals. No data was lost due to technical difficulties over a median monitoring period of 178 minutes per patient. The quality of the study was rated as high (Figure 2).

This study did not report the validation, clinical, or cost outcomes.

Equivital EQ02 Lifemonitor

We found 1 study (N=6) that reported about the Equivital EQ02 Lifemonitor (Hidalgo Ltd) for measuring the HR, RR, ST, and core temperature by using a chest-worn belt monitor (Table 1 and Table 2) [55]. The core temperature was measured using an ingestible pill [77].

Validation outcomes: Acceptable results were found for HR and RR (Figure 2). Skin temperature was outside of the acceptable limits for mean difference and LoA, but the core temperature measurement was considered as acceptable. However, these results were subjected to a high risk of bias at patient selection and reference standard (Figure 2).

This study did not report the feasibility, clinical, and cost outcomes.

Alarm Management System

We found 1 study (N=250) that reported about the Alarm Management System (Covidien; Table 1) [58]. This device was worn at the fingertip and it measures HR and SpO_2 (Table 2).

Feasibility outcomes: The authors reported that 86.6% of the patients completed the monitoring period in the study. Besides, a mean of 4 alarms per week was reported due to decreased SpO_2 in about 75% of the alarms.

Clinical outcomes: The authors reported respiratory event rates, ICU transfer, and RRT calls. However, this occurred 0 times in the control and 1 time in the intervention group. Eventually, the quality of this study was rated as high (Figure 2).

This study did not report about the validation or cost outcomes.

Table 1. Study characteristics.

Author, year	Country	Study design	Setting	Study population	Sample size (N)	Device	Comparison	Conflicts of interest
Prgomet et al, 2016 [39]	Australia	Mixed methods	Single-center hospital	Physicians and nurses of a respiratory and neurosurgery ward	106	ViSi Mobile	None	Not reported
Weller et al, 2017 [40]	USA	Case-control	Single-center hospital	Neurological and neurosurgical patients	736	ViSi Mobile	Manual measurements	None declared
Verillo et al, 2018 [51]	USA	Before-after	Single-center hospital	Orthopedic and trauma patients	422	ViSi Mobile	None	None declared
Weenk et al, 2017 [59]	The Netherlands	Mixed methods	Single-center hospital	Internal and surgical patients	20	ViSi Mobile, HealthPatch	Manual measurements (HR[a], RR[b])	None declared
Watkins et al, 2015 [60]	USA	Cohort	2 hospitals	Nurses	24	ViSi Mobile	None	None declared
Downey et al, 2018a [62]	UK	Pilot Randomized control trial	Single-center hospital	General surgical patients	226	SensiumVitals	Manual and intermittent measurements by nurses (HR, RR, temperature)	None declared
Downey et al, 2018b [56]	UK	Qualitative	Single-center hospital	Surgical patients	12	SensiumVitals	None	None declared
Hernandez-Silveira et al, 2015a [63]	UK	Validation study	Single-center hospital	Surgical and comorbid patients	61	SensiumVitals	Philips Intellivue MP30: 3-lead ECG[c] (HR); Microstream Oridion Capnography (RR)	5 authors were employees of the manufacturing company of the device
Hernandez-Silveira et al, 2015b [61]	UK	Validation study	Laboratory	Healthy subjects	21	SensiumVitals	Rigel 333 patient simulator (HR, RR), Simman (HR), Philips IntelliVue MP30: 2-lead ECG (HR), capnography (RR)	Study was funded by manufacturer, one author was an employee
Downey et al, 2019 [64]	UK	Validation study	Single-center hospital	Major elective surgery patients	51	SensiumVitals	Pulse-oximeter (HR), manually (RR), tympanic thermometer (ST)	None declared
Chan et al, 2013 [65]	USA	Validation study	Laboratory	Healthy subjects	25	HealthPatch	Actiheart, Oridion Capnostream	Authors were employees of the manufacturer of the device
Izmailova et al, 2019 [41]	USA	Validation study	Laboratory	Healthy subjects	6	HealthPatch	Dinamp device (HR), oral thermometer (ST), manual measurement (RR)	None declared
Breteler et al, 2018 [42]	The Netherlands	Validation study	Single-center hospital	Surgical patients	25	HealthPatch	XPREZZON bedside monitor	None declared
Selvaraj et al, 2018 [43]	USA	Validation study	Laboratory	Healthy subjects	57	VitalPatch	Bench testing, Capnostream20, (RR), Actiheart device (HR)	Not reported
Liu et al, 2014 [69]	USA	Validation study	Prehospital	Trauma patients	305	WVSM[d]	LIFEPAK 12 defibrillator/monitor	None declared

Author, year	Country	Study design	Setting	Study population	Sample size (N)	Device	Comparison	Conflicts of interest
Liu et al, 2015 [45]	USA	Cohort	Prehospital	Trauma patients	104	WVSM	None	One author is the CEO[e] of the manufacturing company
Razjouan et al, 2017 [46]	USA	Cohort	Single-center hospital	Hematology and oncology patients	35	Zephyr BioPatch	None	None declared
Boatin et al, 2016 [47]	USA	Mixed methods	Single-center hospital	Full-term pregnant women and nurses	38	Zephyr BioPatch	Pulse-oximeter (HR), manually (RR)	None declared
Kim et al, 2012 [48]	USA	Validation study	Laboratory	Healthy subjects	12	Zephyr BioPatch	12-lead ECG (HR), Model K4 b2, (RR)	None declared
Van Haren et al, 2013 [49]	USA	Cohort	Prehospital	Patients transported by the prehospital provider	113	MiniMedic	LIFEPAK, Propaq MD monitor	None declared
Meisozo et al, 2016 [50]	USA	Validation study	Single-center hospital	Trauma patients in the intensive care unit	59	MiniMedic	GE Solar 8000M multichannel monitor	Not reported
Dur et al, 2019 [52]	USA	Validation study	Laboratory	Healthy subjects	35	Wavelet Wristband	ECG (HR), spirometry sensor (RR), BIOPAC M36	One author was an employee of Wavelet Health
Li et al, 2019 [57]	USA	Validition study	Single-center hospital	Emergency department	17	Biosensor	Capnography (RR)	Two authors were employees of Philips and study was funded by Philips
Ordonnel et al, 2019 [53]	UK	Cohort	Home	Patients with heart failure	13	Proteus patch	None	None declared
Hubner et al, 2015 [54]	Austria	Cohort	Single-center hospital	Patients at the emergency department and nurses who provided care	226	IntelliVue Cableless Measurement Solution	None	None declared
Liu et al, 2013 [55]	China	Validation study	Laboratory	Healthy subjects	6	Equivital EQ02 Lifemonitor	Polar S810i HR Monitor (HR), Spirometer MLT1000L (RR), MLT422/D TSK probe (Temperature)	Not reported
Paul et al, 2019 [58]	Canada	Pilot randomized control trial	Single-center hospital	Mixed surgical patients	250	Covidien Alarm Management System	None	None declared

[a] HR: heart rate.

[b] RR: respiratory rate.

[c] ECG: electrocardiogram.

[d] WVSM: wireless vital signs monitor.

[e] CEO: chief executive officer.

Table 2. Device characteristics.

Device	Manufacturer	Vital signs	Other parameters	Location	BL[a]	CoTy[b]	CR[c] (meter)	EMR[d]	SoA[e]	D[f]	W[g]	S[h]
ViSi Mobile	Sotera Wireless	HR[i], BP[j], RR[k], SpO_2[l], ST[m]	Body posture, fall detection	Upper arm, chest, wrist	14-16 h	Wi-Fi 802.11 radio	180	✓	✓			Clinic
SensiumVitals	Sensium Healthcare	HR, RR, ST	None	Chest, armpit	5 days	Wi-Fi 802.11 b/g	180	✓	✓	✓	✓	Clinic
HealthPatch MD	VitalConnect	HR, RR, ST	HRV[n], fall detection, step count, body posture, R-R interval, stress level, energy expenditure	Chest	3 days	Bluetooth	max. 10		✓	✓	✓	Clinic, home
VitalPatch	VitalConnect	HR, RR, ST	HRV, steps, body posture, fall detection, activity	Chest	5 days	Bluetooth	max. 10	✓	✓	✓	✓	Clinic, home
Wireless Vital Signs Monitor Device	Athena GTX	HR, BP, RR, SpO_2	None	Upper arm, chest, fingertip	7+ h	Wi-Fi 802.11 b/g	180	N/A[o]	✓		✓	Clinic, home
MiniMedic	Athena GTX	HR, SpO_2, ST	PR[p], PWTT[q], Murphy Factor	Forehead, fingertip	12 h	Zigbee 802.15.4	100	N/A	✓		✓	Clinic, home
Zephyr BioPatch	Medtronic	HR, RR, estimated CT[r]	Activity, body posture	Chest	12-28 h	Zephyr ECHO gateway, Bluetooth 2.1+, 3G	N/A	N/A	N/A		N/A	Clinic
Biosensor	Philips	HR, RR, ST	Body posture	Chest	4 days	Bluetooth	Max. 10	✓	✓	✓	✓	Clinic, home
IntelliVue Cableless Measurement Solution	Philips	HR, RR, BP, SpO_2	None	Upper arm, wrist, belly	12-24 h	Short range radio to IntelliVue Guardian Software	<100	✓	✓			Clinic
Wavelet Wristband	Wavelet Health	HR, RR	HRV	Wrist	5 days	Bluetooth	max. 10		N/A		✓	Home
Proteus patch	Proteus Digital Health	HR, RR, ST	None	Upper left chest	7 days	Bluetooth	max. 10	N/A		N/A	✓	Home
EQ02 Lifemonitor	Hidalgo Ltd	HR, RR, ST	ECG[s], accelerometer, body posture, fall detection	Chest with belt	12-48 h	Bluetooth 2.1, 3G, 4G, GPRS[t], CDMA[u]	100	N/A	✓		✓	Clinic, home
Alarm Management System	Covidien	HR, SpO_2	None	Fingertip	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Clinic

[a] BL: battery life.

[b] CoTy: connection type.

[c] CR: connection range.

[d] EMR: electronic medical record.

[e] SoA: system of alerts.

[f] D: disposable.

[g] W: waterproof.

[h] S: setting.

[i] HR: heart rate.

[j] BP: blood pressure.

[k] RR: respiratory rate.

[l] SpO_2: blood oxygen saturation.

[m] ST: skin temperature.

[n] HRV: heart rate variability.

[o] N/A: not applicable.

[p] PR: pulse rate.

[q] PWTT: pulse wave transit time.

[r] CT: core temperature.

[s] ECG: electrocardiogram.

[t] GPRS: general packet radio service.

[u] CDMA: code-division multiple access.

Table 3. Reported outcomes of included studies.

Author, year	Validation outcomes	Feasibility outcomes	Clinical outcomes	Cost outcomes
Prgomet et al, 2016 [39]	—[a]	Knowledge, confidence, perceptions and feedback about continuous monitoring device, interdisciplinary communication regarding deterioration	—	—
Weller et al, 2017 [40]	—	Alarm rate	RRT[b] calls, ICU[c] transfers, unexpected deaths	—
Verillo et al, 2018 [51]	—	Staff satisfaction	Complication rate, RRT calls, ICU transfers, FTR[d] events	—
Weenk et al, 2017 [59]	Bland-Altman agreement	Artifacts, user experiences	—	—
Watkins et al, 2015 [60]	—	Nursing experiences, number of alarms	Log of interventions based on alarms	—
Downey et al, 2018a [62]	—	Patient acceptability and compliance	Time to AB[e], mortality, length of stay, admission to level II or II, 30-day readmission	—
Downey et al, 2018b [56]	—	Patient perceptions	—	—
Hernandez-Silveira et al, 2015a [63]	Bland-Altman agreement	—	—	—
Hernandez-Silveira et al, 2015b [61]	Bland-Altman agreement	—	—	—
Downey et al, 2019 [64]	Bland-Altman agreement	Completeness of continuous patch data	—	—
Chan et al, 2013 [65]	Mean absolute error, root-mean-square error	—	—	—
Izmailova et al, 2019 [41]	Data collection rate, comparison with control, data limitations	Data collection rate, acceptability	—	—
Breteler et al, 2018 [42]	Limits of agreement and bias	Data loss	—	—
Selvaraj et al, 2018 [43]	Bland-Altman agreement	—	—	—
Liu et al, 2014 [69]	—	—	Prediction of life-saving interventions	—
Liu et al, 2015 [45]	—	Percentages of valid measurements and nonzero waveform samples	—	—
Razjouan et al, 2017 [46]	—	Any potential adverse events or complaints as a result of the patch	—	—
Boatin et al, 2016 [47]	Bland-Altman agreement	Acceptability, functionality	—	—
Kim et al, 2012 [48]	Bland-Altman agreement	—	—	—
Van Haren et al, 2013 [49]	Sensitivity, specificity, negative predictive value, positive predictive value, and area under the receiving operating characteristic curves	—	Prediction of life-saving interventions	—
Meisozo et al, 2016 [50]	Paired student *t*-test, Fisher exact tests	—	—	—
Dur et al, 2019 [52]	Pearson correlation coefficients along with Bland-Altman plots and Bland-Altman limits of agreement	—	—	—
Li et al, 2019 [57]	Correlation, mean difference	—	—	—
Ordonnel et al, 2019 [53]	—	Wear-time detection	Sleep detection	—
Hubner et al, 2015 [54]	—	Monitoring time, patient and user experiences	—	—

Author, year	Validation outcomes	Feasibility outcomes	Clinical outcomes	Cost outcomes
Liu et al, 2013 [55]	Bland-Altman agreement, coefficient of variation, ICC[f], SEE[g], Pearson correlation coefficients, ANOVA[h]	—	—	—
Paul et al, 2019 [58]	—	Recruitment rate, acceptance and tolerance, number of alarms per day including type and response, reliability of the system	Respiratory event rate, ICU transfer, RRT calls	—

[a] Not available.

[b] RRT: rapid response time.

[c] ICU: intensive care unit.

[d] FTR: fail-to-rescue.

[e] AB: antibiotic administration.

[f] ICC: intraclass correlation.

[g] SEE: standard error of the estimate.

[h] ANOVA: analysis of variance.

Table 4. Bland-Altman agreement of validation studies.

Device, study, subgroup	HR[a], mean difference (Limits of Agreement)	RR[b], mean difference (Limits of Agreement)	T[c], mean difference (Limits of Agreement)	SpO_2[d], mean difference (Limits of Agreement)	BP syst[e], mean difference (Limits of Agreement)	BP diast[f], mean difference (Limits of Agreement)
VitalPatch, Selvaraj et al, 2018 [43]	0.4 (–8.7/9.5)	–1.8 (–10.1/6.5)	—[g]	—	—	—
HealthPatch, Chan et al, 2013 [65]	—	—	—	—	—	—
HealthPatch, Breteler et al, 2018 [42]	−1.1 (−8.8/6.5)	−2.3 (−15.8/11.2)	—	—	—	—
HealthPatch, Weenk et al, 2017 [59]	–1.52 (–12.55/9.51)	–0.64 (10.32/9.04)	—	—	—	—
ViSi Mobile, Weenk et al, 2017 [59]	–0.2 (–11.06/10.66)	1.19 (–5.53/7.91)	—	0.10 (–3.13/3.33)	0.44 (–23.06/23.94)	–8.00 (–27.46/11.46)
SensiumVitals, Hernandez-Silveira et al, 2015 [63]						
Surgical patients	–0.5 (–3.97/2.97)	0.4 (–6.3/7.1)	—	—	—	—
Cardiovascular disorders (low voltage/variable QRS morphology)	0.97 (–3.73/5.67)	–1.4 (–10.8/8.0)	—	—	—	—
Cardiovascular disorders (atrial fibrillation)	–1.0 (–8.0/6.0)	–1.0 (–9.4/7.0)	—	—	—	—
Metabolic disorders	0.9 (–3.5/5.3)	–0.4 (–11.4/10.6)	—	—	—	—
Diabetes	–0.02 (–6.98/7.02)	0.1 (–7.7/7.9)	—	—	—	—
SensiumVitals, Hernandez-Silveira et al, 2015 [61]	–0.23 (–0.61/0.15)	–0.43 (–6.10/5.20)	—	—	—	—
SensiumVitals, Downey et al, 2019 [64]	1.85 (–23.92/20.22)	2.93 (–8.19/14.05)	0.82 (–1.13/2.78)	—	—	—
Zephyr BioPatch, Boatin et al, 2016 [47][h]	1.6 (–11.6/14.8) - 4.2 (–4.4/22.8)	0.7 (–4.7/6.1) - 4.2 (–1.9/10.3)	0.02 (–1.48/1.52) - 0.5 (–1.3/2.3)	—	—	—
Zephyr BioPatch, Kim et al, 2012 [48]	0.5 (–15.3/16.3)	–0.6 (–5.6/4.4)	—	—	—	—
Wavelet Wristband, Dur et al, 2019 [52]	–0.3 (–2.6/1.9)	1.0 (–3.0/4.0)	—	—	—	—
Biosensor, Li et al, 2019 [57]	—	3.5	—	—	—	—
Equivital EQ02, Liu et al, 2013 [55][i]	1.2 (–5.4/7.8)	0.2 (–2.2/2.6)	0.59 (–0.29/1.47; skin) -0.1 (–0.32/0.12; core)	—	—	—

[a] HR: heart rate.

[b] RR: respiratory rate.

[c] T: temperature.

[d] SpO_2: oxygen saturation.

[e] BP syst: systolic blood pressure.

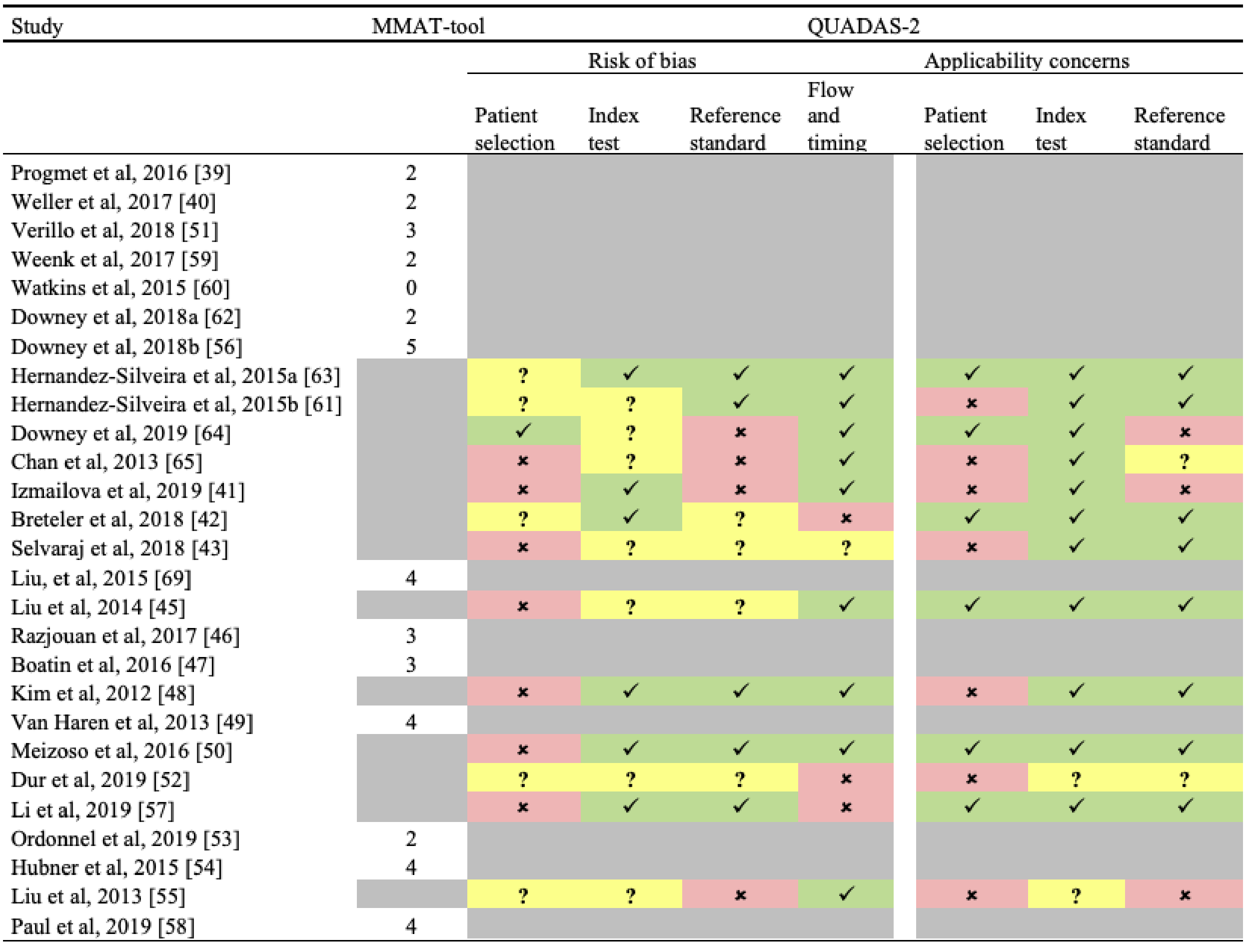
[f] BP diast: diastolic blood pressure.

[g] Not available.

[h] This study reported the 25th and 75th percentile.

[i] This study reported the Bland-Altman agreement of two types of temperature: skin and core temperature.

Figure 2. Quality assessment of the included studies. Check marks: low risk of bias; Crosses: high risk of bias; Question marks: unclear risk of bias; Grey cells: Quality assessment tool not used for the study.



Study	MMAT-tool	QUADAS-2						
		Risk of bias				Applicability concerns		
		Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Progmet et al, 2016 [39]	2							
Weller et al, 2017 [40]	2							
Verillo et al, 2018 [51]	3							
Weenk et al, 2017 [59]	2							
Watkins et al, 2015 [60]	0							
Downey et al, 2018a [62]	2							
Downey et al, 2018b [56]	5							
Hernandez-Silveira et al, 2015a [63]		?	✓	✓	✓	✓	✓	✓
Hernandez-Silveira et al, 2015b [61]		?	?	✓	✓	×	✓	✓
Downey et al, 2019 [64]		✓	?	×	✓	✓	✓	×
Chan et al, 2013 [65]		×	?	×	✓	×	✓	?
Izmailova et al, 2019 [41]		×	✓	×	✓	×	✓	×
Breteler et al, 2018 [42]		?	✓	?	×	✓	✓	✓
Selvaraj et al, 2018 [43]		×	?	?	?	×	✓	✓
Liu, et al, 2015 [69]	4							
Liu et al, 2014 [45]		×	?	?	✓	✓	✓	✓
Razjouan et al, 2017 [46]	3							
Boatin et al, 2016 [47]	3							
Kim et al, 2012 [48]		×	✓	✓	✓	×	✓	✓
Van Haren et al, 2013 [49]	4							
Meizoso et al, 2016 [50]		×	✓	✓	✓	✓	✓	✓
Dur et al, 2019 [52]		?	?	?	×	×	?	?
Li et al, 2019 [57]		×	✓	✓	×	✓	✓	✓
Ordonnel et al, 2019 [53]	2							
Hubner et al, 2015 [54]	4							
Liu et al, 2013 [55]		?	?	×	✓	×	?	×
Paul et al, 2019 [58]	4							

Discussion

Summary of Evidence

In this study, we aimed to provide a systematic review of the current evidence on wearable wireless continuous monitoring devices for vital signs monitoring. We included 27 studies, which evaluated 13 different wearable devices. Overall, the studies predominantly evaluated the validation of the recorded data (N=15) or the feasibility (N=15) of these devices. Clinical outcomes were only reported in 6 studies, and studies describing the cost outcomes are still lacking. Although 13 different devices were included in this review, these devices did not share the same indication in terms of monitoring. In general, 2 main target indications could be identified. First, the ViSi Mobile, WVSM Device, MiniMedic, and IntelliVue Cableless Measurement Solution were designed for more extensive prehospital (ambulance) or clinical physiological monitoring. This monitoring level may be comparable to standard ICU monitoring, and therefore, these devices are usually bulkier wearable devices. Second, patch, wristband, and harness devices such as the SensiumVitals, VitalPatch, Philips Biosensor, Zephyr BioPatch, EQ02 Lifemonitor, Alarm Management System, Wavelet Wristband, and the Proteus patch were designed for ambulant wireless clinical monitoring of only a few basic vital signs. These devices are possibly more suitable for patients in the general ward and for monitoring the vital signs at home.

Regarding the validation of the devices, a few considerations should be taken into account. Many of these studies were conducted in healthy volunteers, which may introduce a bias owing to the lack of deviating vital signs values when compared to the vital signs of the actual patients. Further, for technical reasons, vital signs cannot be measured continuously by wearable sensors with equal accuracy. In particular, the RR and temperatures still appear to be difficult to be measured reliably in several included studies. In fact, the optimal reference standard for measuring RR has still not been found, although it is considered to be the most important parameter for predicting clinical deterioration [78-81]. In addition, the optimal method for measuring temperature by using wearable wireless devices has yet to be found. Most devices measure the skin temperature, which is known to be unreliable as equivalent for core temperature [82-84].

Feasibility outcomes were focused on acceptability by health care professionals and patients. In general, both groups were positive about the deployment of the devices. In addition, the operation of the system was evaluated, such as the completeness of the measurements and the number and appropriateness of the alarms. Both outcomes were assessed as feasible.

The impact of these devices on clinical outcomes is still unclear because most included studies were underpowered to demonstrate any significant effect. However, multiple studies described cases wherein a complication was recognized earlier by the device and acted upon in a timely manner.

Regarding costs, no outcomes were reported about the devices in the included studies. Such data may however be essential for preparing future business cases for large-scale implementation, considering the relatively high cost of such monitoring devices and platforms [85].

Previously published reviews on continuous monitoring did not focus on wearable devices, except for one, but this was not a systematic review [32]. We found comparable but also contrasting results in that study [32]. The review of Joshi et al [32] reported the same devices as those reported by us as well as some other devices that we excluded since there were no published studies about those devices or they were published before 2009. In line with our results, they also concluded that the diagnostic accuracy of the devices was suboptimal, especially the alarm rates and the false alarms. In addition, they also indicated that there were no sufficiently powered studies to show beneficial clinical effects or cost-effectiveness.

In a review of nonwearable devices, Cardona-Morrell et al [14] found that early detection of deterioration was enhanced but there were no significant improvements in the clinical outcomes, which is in line with our findings regarding wearable devices. This could be explained by the heterogeneous and underpowered character of the included studies [14]. Downey et al [86] also came to this conclusion and further stated that continuous monitoring seems to be feasible in terms of the frequency of implementation in hospitals; they found that patient and nurse perceptions were positive and that continuous monitoring may be cost-efficient.

Limitations

This systematic review had several limitations. First, the quality varied across the included studies. Several accuracy studies contained high risk of bias regarding patient selection as well as the applicability. Further, the reference standard was often not free from potential bias. Considering the studies assessed with the MMAT tool, quality was predominantly rated 2 or 3 out of 5; therefore, bias is present. Moreover, assessing the quality of the studies and comparing these studies was difficult owing to the heterogeneity of the included studies. Therefore, performing a meta-analysis was not possible owing to the heterogeneity in the devices and the outcomes. Second, 5 of the included studies had possible conflicts of interest owing to funding by the manufacturer or because employees of the manufacturing companies of the devices played a role in the conduct of the study. This highlights the possible risk of reporting and publication bias within this field of research. Third, there were some limitations about the search. We only focused on devices that measured at least two vital signs. However, this cut-off was based on previous studies about the predictive value for clinical deterioration. These studies found that the more vital signs are monitored, the more accurate the detection is [87,88]. Besides, we only focused on off-the-shelf devices with a clearance by the CE mark or FDA as a medical device for clinical use. We excluded 42 prototype studies that were considered to be less clinically relevant for health care professionals. However, this indicates that there may be many more monitoring devices that will be launched in the health care market in the future. Besides, the review was restricted to English and Dutch publications published from 2009 and after. Only a few studies were excluded based on language and the older studies were considered be less clinically relevant owing to outdated technology. Fourth, we prespecified the clinically relevant mean difference and LoA for vital signs. It may be clinically desirable to redefine acceptable accuracy limits depending on the value of the vital signs measured and the patient population. For example, a difference of 3 breaths per minute is more clinically relevant in a range of 5-8 breaths per minute than with 30-33 breaths per minute. However, reliable evidence or guidelines for continuous monitoring of vital signs are currently lacking.

Clinical Implications

This review outlines several important clinical implications before health systems may proceed to large-scale implementation of wearable wireless continuous monitoring devices for vital signs monitoring for patients in the hospital and at home. For both settings, vital signs data measurements should be accurate, reliable, and validated in clinical studies. This is especially important for the home setting, wherein a health care professional is not readily available to assess the clinical condition of the patient. For further optimization, the monitoring measurements should preferably be incorporated into an early warning score system supported by a validated decision support algorithm [89]. These analysis algorithms should be further enhanced to prevent too many alarms in order to avoid alarm fatigue [90]. Further, for optimal adoption into clinical workflows, the vital signs measurements should preferably be integrated into the electronic medical record. This will likely improve commitment and compliance from nurses and doctors and will also allow for the summarized monitoring data to be archived in the patient records [32]. When all such factors are optimized, it is anticipated that studies will be able to show a significant effect on clinical outcomes. For monitoring patients at home, the patient data need to be sent to health care professionals through a stable and secure wireless connection. Such a system will need to be embedded in a validated care work flow, thereby providing alarm reviews by care professionals who will assess, make an initial phone call, and then escalate to a home visit by a nurse or direct the patient to the emergency department when needed [91]. Furthermore, for home monitoring, the devices should be small, flexible, and hypoallergenic and not bother patients during their daily activities [18,24]. Battery life, which currently ranges from 3 to 7 days in most devices, may be further extended especially for long-term monitoring of patients with chronic diseases such as heart failure [18,19]. Eventually, when all the conditions are optimized, larger studies may be able to demonstrate that continuous home monitoring safely allows for routine early discharge from the hospital. Further, such a system may potentially provide timely detection of complications, and thereby prevent readmissions, improve overall outcomes, and decrease health care costs [21,92].

Conclusions

Continuous monitoring devices are mostly still in the validation and feasibility phases. Besides, studies reporting clinical outcomes are still sparse and cost outcome studies are still lacking. Such studies are needed to help health care professionals and administrators in their decision making regarding the implementation of these devices on a large scale in clinical practice or in home monitoring.

Acknowledgments

This study was funded by the Innovation and Science Fund of Isala (grant number: INNO 1915).

Authors' Contributions

JL conceived the presented idea. JL, CH, HvW, LS, and GP designed the methods. JL conducted the search, JL and CH independently included the studies, extracted the data, and were responsible for the quality assessment whereas HvW, LS, and GP provided supervision during all the stages. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string of PubMed/MEDLINE database.
[DOCX File , 12 KB - jmir_v22i6e18636_app1.docx]

References

1. No authors listed. Guidelines for intensive care unit admission, discharge, and triage. Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine. Crit Care Med 1999 Mar;27(3):633-638. [Medline: 10199547]
2. Padilla RM, Mayo A. Clinical deterioration: A concept analysis. J Clin Nurs 2018 Feb 15;27(7-8):1360-1368. [doi: 10.1111/jocn.14238] [Medline: 29266536]
3. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: The impact of outcome selection. Resuscitation 2013 May;84(5):564-568. [doi: 10.1016/j.resuscitation.2012.09.024] [Medline: 23022075]
4. Downey C, Tahir W, Randell R, Brown J, Jayne D. Strengths and limitations of early warning scores: A systematic review and narrative synthesis. International Journal of Nursing Studies 2017 Nov;76:106-119. [doi: 10.1016/j.ijnurstu.2017.09.003] [Medline: 28950188]
5. Clifton L, Clifton DA, Pimentel MAF, Watkinson PJ, Tarassenko L. Predictive Monitoring of Mobile Patients by Combining Clinical Observations With Data From Wearable Sensors. IEEE J. Biomed. Health Inform 2014 May;18(3):722-730. [doi: 10.1109/jbhi.2013.2293059]
6. Credland N, Dyson J, Johnson MJ. What are the patterns of compliance with Early Warning Track and Trigger Tools: A narrative review. Applied Nursing Research 2018 Dec;44:39-47. [doi: 10.1016/j.apnr.2018.09.002] [Medline: 30389058]
7. Perman SM, Stanton E, Soar J, Berg RA, Donnino MW, Mikkelsen ME, et al. Location of In‐Hospital Cardiac Arrest in the United States—Variability in Event Rate and Outcomes. JAHA 2016 Oct 03;5(10):1-7. [doi: 10.1161/jaha.116.003638] [Medline: 27688235]
8. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008 Jun;17(3):216-223 [FREE Full text] [doi: 10.1136/qshc.2007.023622] [Medline: 18519629]
9. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in Older Adults: Evidence for a Phenotype. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 2001 Mar 01;56(3):M146-M157. [doi: 10.1093/gerona/56.3.M146] [Medline: 11253156]
10. Maurits EEM, de Veer AJE, Francke AL. Ervaringen van verpleegkundigen, verzorgenden, begeleiders en praktijkondersteuners. NIVEL. 2016 Jun 09. Inspelen op veranderingen in de zorg URL: https://www.nivel.nl/nl/publicatie/inspelen-op-veranderingen-de-zorg-ervaringen-van-verpleegkundigen-verzorgenden [accessed 2019-08-03]
11. Benjamin E, Virani S. Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. Circulation 2018 Mar 20;137(12):e67-e492. [doi: 10.1161/CIR.0000000000000558] [Medline: 29386200]
12. Nolan JP, Soar J, Smith GB, Gwinnutt C, Parrott F, Power S, et al. Incidence and outcome of in-hospital cardiac arrest in the United Kingdom National Cardiac Arrest Audit. Resuscitation 2014 Aug;85(8):987-992. [doi: 10.1016/j.resuscitation.2014.04.002] [Medline: 29386200]
13. Pearse RM, Moreno RP, Bauer P, Pelosi P, Metnitz P, Spies C, et al. Mortality after surgery in Europe: a 7 day cohort study. The Lancet 2012 Sep;380(9847):1059-1065. [doi: 10.1016/s0140-6736(12)61148-9]

14. Cardona-Morrell M, Prgomet M, Turner RM, Nicholson M, Hillman K. Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. Int J Clin Pract 2016 Oct;70(10):806-824. [doi: 10.1111/ijcp.12846] [Medline: 27582503]
15. Jones D, Mitchell I, Hillman K, Story D. Defining clinical deterioration. Resuscitation 2013 Aug;84(8):1029-1034. [doi: 10.1016/j.resuscitation.2013.01.013] [Medline: 23376502]
16. Elliott M, Coventry A. Critical care: the eight vital signs of patient monitoring. Br J Nurs 2012;21(10):621-625. [doi: 10.12968/bjon.2012.21.10.621] [Medline: 22875303]
17. Appelboom G, Camacho E, Abraham ME, Bruce SS, Dumont EL, Zacharia BE, et al. Smart wearable body sensors for patient self-assessment and monitoring. Arch Public Health 2014 Aug;72(1):28 [FREE Full text] [doi: 10.1186/2049-3258-72-28] [Medline: 25232478]
18. Darwish A, Hassanien AE. Wearable and Implantable Wireless Sensor Network Solutions for Healthcare Monitoring. Sensors 2011 May 26;11(6):5561-5595. [doi: 10.3390/s110605561]
19. Sahandi R, Noroozi S, Roushan G, Heaslip V, Liu Y. Wireless technology in the evolution of patient monitoring on general hospital wards. J Med Eng Technol 2010 Jan;34(1):51-63. [doi: 10.3109/03091900903336902] [Medline: 19929237]
20. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
21. Bellomo R, Ackerman M, Bailey M, Beale R, Clancy G, Danesh V, et al. A controlled trial of electronic automated advisory vital signs monitoring in general hospital wards*. Critical Care Medicine 2012;40(8):2349-2361. [doi: 10.1097/ccm.0b013e318255d9a0]
22. Schmidt PE, Meredith P, Prytherch DR, Watson D, Watson V, Killen RM, et al. Impact of introducing an electronic physiological surveillance system on hospital mortality. BMJ Qual Saf 2015 Jan;24(1):10-20. [doi: 10.1136/bmjqs-2014-003073] [Medline: 25249636]
23. Kollef MH, Heard K, Chen Y, Lu C, Martin N, Bailey T. Mortality and Length of Stay Trends Following Implementation of a Rapid Response System and Real-Time Automated Clinical Deterioration Alerts. Am J Med Qual 2017;32(1):12-18 [FREE Full text] [doi: 10.1177/1062860615613841] [Medline: 26566998]
24. Majumder S, Mondal T, Deen M. Wearable Sensors for Remote Health Monitoring. Sensors 2017 Jan 12;17(1):130. [doi: 10.3390/s17010130]
25. Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. Crit Care 2019 May 30;23(1):194 [FREE Full text] [doi: 10.1186/s13054-019-2485-7] [Medline: 31146792]
26. Michard F, Gan T, Kehlet H. Digital innovations and emerging technologies for enhanced recovery programmes. Br J Anaesth 2017 Jul 01;119(1):31-39 [FREE Full text] [doi: 10.1093/bja/aex140] [Medline: 28605474]
27. Fung E, Järvelin MR, Doshi RN, Shinbane JS, Carlson SK, Grazette LP, et al. Electrocardiographic patch devices and contemporary wireless cardiac monitoring. Front Physiol 2015;6:149 [FREE Full text] [doi: 10.3389/fphys.2015.00149] [Medline: 26074823]
28. Iqbal MH, Aydin A, Brunckhorst O, Dasgupta P, Ahmed K. A review of wearable technology in medicine. J R Soc Med 2016 Oct;109(10):372-380 [FREE Full text] [doi: 10.1177/0141076816663560] [Medline: 27729595]
29. Paton C, Hansen M, Fernandez-Luque L, Lau AYS. Self-Tracking, Social Media and Personal Health Records for Patient Empowered Self-Care. Yearb Med Inform 2018 Mar 10;21(01):16-24. [doi: 10.1055/s-0038-1639425]
30. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
31. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions 6th version (updated July 2019). Cochrane. 2019. URL: https://training.cochrane.org/handbook/current [accessed 2019-08-01]
32. Joshi M, Ashrafian H, Aufegger L, Khan S, Arora S, Cooke G, et al. Wearable sensors to improve detection of patient deterioration. Expert Rev Med Devices 2019 Feb;16(2):145-154. [doi: 10.1080/17434440.2019.1563480] [Medline: 30580650]
33. Harford M, Catherall J, Gerry S, Young J, Watkinson P. Availability and performance of image-based, non-contact methods of monitoring heart rate, blood pressure, respiratory rate, and oxygen saturation: a systematic review. Physiol Meas 2019 Jul 03;40(6):06TR01. [doi: 10.1088/1361-6579/ab1f1d] [Medline: 31051494]
34. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
35. Henrikson N, Skelly A. Economic Studies Part 2: Evaluating the Quality. Evidence-Based Spine-Care Journal 2013 May 1;04(01):002-005. [doi: 10.1055/s-0033-1341594]
36. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Mixed Methods Appraisal Tool (MMAT), version 2018, Registration of Copyright (#1148552).: Canadian Intellectual Property Office, Industry Canada; 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127425851/MMAT_2018_criteria-manual_2018-04-04.pdf [accessed 2019-08-04]

37. Whiting PF. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med 2011 Oct 18;155(8):529. [doi: 10.7326/0003-4819-155-8-201110180-00009]
38. Ofman JJ, Sullivan SD, Neumann PJ, Chiou C, Henning JM, Wade SW, et al. Examining the Value and Quality of Health Economic Analyses: Implications of Utilizing the QHES. JMCP 2003 Jan;9(1):53-61. [doi: 10.18553/jmcp.2003.9.1.53]
39. Prgomet M, Cardona-Morrell M, Nicholson M, Lake R, Long J, Westbrook J, et al. Vital signs monitoring on general wards: clinical staff perceptions of current practices and the planned introduction of continuous monitoring technology. Int J Qual Health Care 2016 Jun 17;28(4):515-521. [doi: 10.1093/intqhc/mzw062]
40. Weller RS, Foard KL, Harwood TN. Evaluation of a wireless, portable, wearable multi-parameter vital signs monitor in hospitalized neurological and neurosurgical patients. J Clin Monit Comput 2018 Oct;32(5):945-951. [doi: 10.1007/s10877-017-0085-0] [Medline: 29214598]
41. Izmailova ES, McLean IL, Bhatia G, Hather G, Cantor M, Merberg D, et al. Evaluation of Wearable Digital Devices in a Phase I Clinical Trial. Clin Transl Sci 2019 Jan 11;12(3):247-256. [doi: 10.1111/cts.12602]
42. Breteler MJM, Huizinga E, van Loon K, Leenen LPH, Dohmen DAJ, Kalkman CJ, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. BMJ Open 2018 Feb 27;8(2):e020162. [doi: 10.1136/bmjopen-2017-020162]
43. Selvaraj N, Nallathambi G, Moghadam R, Aga A. Fully Disposable Wireless Patch Sensor for Continuous Remote Patient Monitoring. Conf Proc IEEE Eng Med Biol Soc 2018 Jul;2018:1632-1635. [doi: 10.1109/EMBC.2018.8512569] [Medline: 30440706]
44. Liu NT, Holcomb JB, Wade CE, Darrah MI, Salinas J. Evaluation of standard versus nonstandard vital signs monitors in the prehospital and emergency departments. Journal of Trauma and Acute Care Surgery 2014;77:S121-S126. [doi: 10.1097/ta.0000000000000192]
45. Liu NT, Holcomb JB, Wade CE, Darrah MI, Salinas J. Data quality of a wearable vital signs monitor in the pre-hospital and emergency departments for enhancing prediction of needs for life-saving interventions in trauma patients. Journal of Medical Engineering & Technology 2015 Jun 19;39(6):316-321. [doi: 10.3109/03091902.2015.1054524]
46. Razjouyan J, Grewal GS, Rishel C, Parthasarathy S, Mohler J, Najafi B. Activity Monitoring and Heart Rate Variability as Indicators of Fall Risk: Proof-of-Concept for Application of Wearable Sensors in the Acute Care Setting. J Gerontol Nurs 2017 Mar 02;43(07):53-62. [doi: 10.3928/00989134-20170223-01]
47. Boatin A, Wylie B, Goldfarb I, Azevedo R, Pittel E, Ng C, et al. Wireless Vital Sign Monitoring in Pregnant Women: A Functionality and Acceptability Study. Telemed J E Health 2016 Jul;22(7):564-571. [doi: 10.1089/tmj.2015.0173] [Medline: 27172448]
48. Kim J, Roberge R, Powell J, Shafer A, Jon Williams W. Measurement Accuracy of Heart Rate and Respiratory Rate during Graded Exercise and Sustained Exercise in the Heat Using the Zephyr BioHarnessTM. Int J Sports Med 2012 Nov 22;34(06):497-501. [doi: 10.1055/s-0032-1327661]
49. Van Haren RM, Thorson CM, Valle EJ, Busko AM, Jouria JM, Livingstone AS, et al. Novel prehospital monitor with injury acuity alarm to identify trauma patients who require lifesaving intervention. Journal of Trauma and Acute Care Surgery 2014;76(3):743-749. [doi: 10.1097/ta.0000000000000099]
50. Meizoso JP, Allen CJ, Ray JJ, Van Haren RM, Teisch LF, Baez XR, et al. Evaluation of Miniature Wireless Vital Signs Monitor in a Trauma Intensive Care Unit. Military Medicine 2016 May;181(5S):199-204. [doi: 10.7205/milmed-d-15-00162]
51. Verrillo SC, Cvach M, Hudson KW, Winters BD. Using Continuous Vital Sign Monitoring to Detect Early Deterioration in Adult Postoperative Inpatients. Journal of Nursing Care Quality 2019;34(2):107-113. [doi: 10.1097/ncq.0000000000000350]
52. Dur O, Rhoades C, Ng MS, Elsayed R, van Mourik R, Majmudar MD. Design Rationale and Performance Evaluation of the Wavelet Health Wristband: Benchtop Validation of a Wrist-Worn Physiological Signal Recorder. JMIR Mhealth Uhealth 2018 Oct 16;6(10):e11040 [FREE Full text] [doi: 10.2196/11040] [Medline: 30327288]
53. OrDonnell J, Velardo C, Shah SA, Khorshidi GS, Salvi D, Rahimi K, et al. Physical Activity and Sleep Analysis of Heart Failure Patients using Multi-sensor Patches. Conf Proc IEEE Eng Med Biol Soc 2018 Jul:6092-6095. [doi: 10.1109/embc.2018.8513594] [Medline: 30441725]
54. Hubner P, Schober A, Sterz F, Stratil P, Wallmueller C, Testori C, et al. Surveillance of Patients in the Waiting Area of the Department of Emergency Medicine. Medicine 2015;94(51):e2322. [doi: 10.1097/md.0000000000002322]
55. Liu Y, Zhu SH, Wang GH, Ye F, Li PZ. Validity and Reliability of Multiparameter Physiological Measurements Recorded by the Equivital Lifemonitor During Activities of Various Intensities. Journal of Occupational and Environmental Hygiene 2013 Feb;10(2):78-85. [doi: 10.1080/15459624.2012.747404]
56. Downey C, Brown J, Jayne D, Randell R. Patient attitudes towards remote continuous vital signs monitoring on general surgery wards: An interview study. International Journal of Medical Informatics 2018 Jun;114:52-56. [doi: 10.1016/j.ijmedinf.2018.03.014]
57. Li T, Divatia S, McKittrick J, Moss J, Hijnen NM, Becker LB. A pilot study of respiratory rate derived from a wearable biosensor compared with capnography in emergency department patients. OAEM 2019 May;Volume 11:103-108. [doi: 10.2147/oaem.s198842]

58. Paul JE, Chong MA, Buckley N, Harsha P, Shanthanna H, Tidy A, et al. Vital sign monitoring with continuous pulse oximetry and wireless clinical notification after surgery (the VIGILANCE pilot study)-a randomized controlled pilot trial. Pilot Feasibility Stud 2019 Feb 26;5(1):36 [FREE Full text] [doi: 10.1186/s40814-019-0415-8] [Medline: 30858986]
59. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study. JMIR Mhealth Uhealth 2017 Jul 05;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
60. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. J Clin Nurs 2016 Jan;25(1-2):278-281 [FREE Full text] [doi: 10.1111/jocn.13102] [Medline: 26531215]
61. Hernandez-Silveira M, Wieczorkowski-Rettinger K, Ang S, Burdett A. Preliminary assessment of the SensiumVitals®: A low-cost wireless solution for patient surveillance in the general wards. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:4931-4937. [doi: 10.1109/EMBC.2015.7319498] [Medline: 26737398]
62. Downey C, Randell R, Brown J, Jayne DG. Continuous Versus Intermittent Vital Signs Monitoring Using a Wearable, Wireless Patch in Patients Admitted to Surgical Wards: Pilot Cluster Randomized Controlled Trial. J Med Internet Res 2018 Dec 11;20(12):e10802 [FREE Full text] [doi: 10.2196/10802] [Medline: 30538086]
63. Hernandez-Silveira M, Ahmed K, Ang S, Zandari F, Mehta T, Weir R, et al. Assessment of the feasibility of an ultra-low power, wireless digital patch for the continuous ambulatory monitoring of vital signs. BMJ Open 2015 May 19;5(5):e006606 [FREE Full text] [doi: 10.1136/bmjopen-2014-006606] [Medline: 25991447]
64. Downey C, Ng S, Jayne D, Wong D. Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: a clinical validation study from the TRaCINg trial. BMJ Open 2019 Aug 15;9(8):e031150 [FREE Full text] [doi: 10.1136/bmjopen-2019-031150] [Medline: 31420399]
65. Chan AM, Selvaraj N, Ferdosi N, Narasimhan R. Wireless patch sensor for remote monitoring of heart rate, respiration, activity, and falls. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2013:6115-6118. [doi: 10.1109/embc.2013.6610948]
66. Sotera W. Sotera Wireless. ViSi Mobile System Technical Specifications URL: https://medaval.ie/docs/specs/Sotera-VisiMobile-Specs.pdf [accessed 2019-08-24]
67. Sensium Healthcare. Early detection of patient deterioration. 2019. Sensium® URL: https://www.sensium.co.uk/_assets/media/documents/brochures/brochure.pdf [accessed 2019-11-19]
68. VitalConnect. The VitalConnect Solution. 2017. The VitalConnect Solution: shattering the paradigm of patient care as we know it today URL: https://vitalconnect.com/wp-content/uploads/2017/01/MKT-034_A_VitalConnect-Product-Brochure.pdf [accessed 2019-11-04]
69. Liu NT, Holcomb JB, Wade CE, Darrah MI, Salinas J. Evaluation of standard versus nonstandard vital signs monitors in the prehospital and emergency departments. Journal of Trauma and Acute Care Surgery 2014;77:S121-S126. [doi: 10.1097/ta.0000000000000192]
70. Athena GTX. Published. 2019. WVSM®: Multi patient monitoring URL: https://athenagtx.com/wp-content/uploads/2018/02/WVSM_Brochure_RevF.pdf [accessed 2019-10-26]
71. Athena GTX. Product catalog. mini-MEDIC® URL: https://healthmanagement.org/products/view/ecg-patient-monitor-infant-emergency-wireless-mini-medic-r-athena-gtx [accessed 2019-10-27]
72. Medtronic. Medtronic. Zephyr BioPatchTM Monitoring Device for Human Performance (HP) URL: https://www.zephyranywhere.com/media/download/zephyr-performance-biopatch-hp-brochure.pdf [accessed 2019-10-16]
73. Philips. Published. 2019. Wearable Biosensor URL: https://www.philips.nl/healthcare/product/HC989803196871/wearable-biosensor-wireless-remote-sensing-device [accessed 2019-10-22]
74. Wavelet H. Published. 2019. Wavelet wristband and pod URL: https://usermanual.wiki/Amiigo/W1002/html [accessed 2019-10-27]
75. Proteus Digital Health, Inc. Proteus Digital Health Feedback Device. 2019. URL: https://tinyurl.com/y9av6kso [accessed 2019-10-22]
76. Philips Healthcare. Published. 2013. Small, lightweight, and cableless URL: https://philipsproductcontent.blob.core.windows.net/assets/20171012/ba825e4be9b348e0a383a80a001d601f.pdf [accessed 2019-10-22]
77. Hildago Ltd.. Published. 2012. EQ02 LifeMonitor URL: https://www.equivital.com/assets/common/SEM_Data_Sheet_General_HIDA3330-DSG-02.2_.2_.pdf [accessed 2019-10-16]
78. Mochizuki K, Shintani R, Mori K, Sato T, Sakaguchi O, Takeshige K, et al. Importance of respiratory rate for the prediction of clinical deterioration after emergency department discharge: a single-center, case-control study. Acute Med Surg 2017 Apr;4(2):172-178 [FREE Full text] [doi: 10.1002/ams2.252] [Medline: 29123857]
79. Flenady T, Dwyer T, Applegarth J. Accurate respiratory rates count: So should you!. Australas Emerg Nurs J 2017 Feb;20(1):45-47. [doi: 10.1016/j.aenj.2016.12.003] [Medline: 28073649]
80. Parkes R. Rate of respiration: the forgotten vital sign. Emerg Nurse 2011 May;19(2):12-7; quiz 18. [doi: 10.7748/en2011.05.19.2.12.c8504] [Medline: 21675463]

81. Yonge JD, Bohan PK, Watson JJ, Connelly CR, Eastes L, Schreiber MA. The Respiratory Rate: A Neglected Triage Tool for Pre-hospital Identification of Trauma Patients. World J Surg 2018 May;42(5):1321-1326. [doi: 10.1007/s00268-017-4353-4] [Medline: 29214444]
82. Mendt S, Maggioni MA, Nordine M, Steinach M, Opatz O, Belavý D, et al. Circadian rhythms in bed rest: Monitoring core body temperature via heat-flux approach is superior to skin surface temperature. Chronobiol Int 2017;34(5):666-676. [doi: 10.1080/07420528.2016.1224241] [Medline: 27726448]
83. Henriksson E, Lamia KA. Adipose Clocks: Burning the Midnight Oil. J Biol Rhythms 2015 Oct;30(5):364-373 [FREE Full text] [doi: 10.1177/0748730415581234] [Medline: 25926681]
84. Albrecht U. Timing to perfection: the biology of central and peripheral circadian clocks. Neuron 2012 Apr 26;74(2):246-260 [FREE Full text] [doi: 10.1016/j.neuron.2012.04.006] [Medline: 22542179]
85. Slight SP, Franz C, Olugbile M, Brown HV, Bates DW, Zimlichman E. The Return on Investment of Implementing a Continuous Monitoring System in General Medical-Surgical Units*. Critical Care Medicine 2014;42(8):1862-1868. [doi: 10.1097/ccm.0000000000000340]
86. Downey C, Chapman S, Randell R, Brown J, Jayne D. The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. Int J Nurs Stud 2018 Aug;84:19-27. [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
87. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. Resuscitation 2016 May;102:1-5 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.02.005] [Medline: 26898412]
88. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D. Early detection of impending physiologic deterioration among patients who are not in intensive care: development of predictive models using data from an automated electronic medical record. J Hosp Med 2012 Mar;7(5):388-395. [doi: 10.1002/jhm.1929] [Medline: 22447632]
89. Alam N, Hobbelink E, van Tienhoven A, van de Ven P, Jansma E, Nanayakkara P. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. Resuscitation 2014 May;85(5):587-594. [doi: 10.1016/j.resuscitation.2014.01.013] [Medline: 24467882]
90. Sendelbach S, Funk M. Alarm Fatigue. AACN Advanced Critical Care 2013;24(4):378-386. [doi: 10.1097/nci.0b013e3182a903f9]
91. Kowalski R, Capan M, Lodato P, Mosby D, Thomas T, Arnold R, et al. Optimizing usability and signal capture: a proactive risk assessment for the implementation of a wireless vital sign monitoring system. J Med Eng Technol 2017 Nov;41(8):623-629. [doi: 10.1080/03091902.2017.1382589] [Medline: 29027496]
92. Frost&Sullivan. Topline Finding Top-Line Opportunities in Bottom- line HC Market. Finding Top-Line Opportunities in a Bottom-Line Healthcare Market: Revenue Generation Opportunities Created by an Innovative Contact-Free Continuous Monitoring Solution URL: http://frost-apac.com/BDS/whitepaper/Whitepaper%20Topline%20Finding%20Top-Line%20Opportunities%20in%20Bottom-%20line%20HC%20Market.pdf [accessed 2019-12-02]

Abbreviations

BP: blood pressure
CE: Conformité Européenne
ECG: electrocardiogram
FDA: Food and Drug Administration
FTR: fail-to-rescue
HR: heart rate
ICU: intensive care unit
LoA: limits of agreement
LSI: life-saving intervention
MMAT: mixed methods appraisal tool
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QHES: quality of health economic studies
QUADAS-2: quality assessment of diagnostic accuracy studies 2nd edition
RR: respiratory rate
RRT: rapid response team
SpO2: blood oxygen saturation
WVSM: wireless vital signs monitor

Edited by G Eysenbach; submitted 09.03.20; peer-reviewed by K Marko, C Fincham, B Johnson; comments to author 31.03.20; revised version received 07.04.20; accepted 14.05.20; published 17.06.20.

Please cite as:
Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA
Current Evidence for Continuous Vital Signs Monitoring by Wearable Wireless Devices in Hospitalized Adults: Systematic Review
J Med Internet Res 2020;22(6):e18636
URL: http://www.jmir.org/2020/6/e18636/
doi:10.2196/18636
PMID:32469323

Review

Engaging Children and Young People in Digital Mental Health Interventions: Systematic Review of Modes of Delivery, Facilitators, and Barriers

Shaun Liverpool[1], MSc; Catarina Pinheiro Mota[2,3], PhD; Célia M D Sales[4], PhD; Anja Čuš[5], MSc; Sara Carletto[6], PhD; Camellia Hancheva[7], PhD; Sónia Sousa[8], PhD; Sonia Conejo Cerón[9], PhD; Patricia Moreno-Peral[9], PhD; Giada Pietrabissa[10,11], PhD; Bettina Moltrecht[1], PhD; Randi Ulberg[12], PhD; Nuno Ferreira[13], PhD; Julian Edbrooke-Childs[1], PhD

[1]Evidence-Based Practice Unit, University College London and Anna Freud National Centre for Children and Families, London, United Kingdom
[2]Center for Psychology, University of Porto, Porto, Portugal
[3]University of Trás-os-Montes and Alto Douro, Porto, Portugal
[4]Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal
[5]Department of Child and Adolescent Psychiatry, Medical University of Vienna, Vienna, Austria
[6]Department of Clinical and Biological Sciences, University of Turin, Turin, Italy
[7]Faculty of Philosophy, General, Experimental, Developmental, and Health Psychology, Sofia University, Sofia, Bulgaria
[8]School of Digital Technologies, Tallinn University, Tallinn, Estonia
[9]Biomedical Research Institute of Malaga, Málaga, Spain
[10]Department of Psychology, Catholic University of Milan, Milan, Italy
[11]Psychology Research Laboratory, IRCCS Istituto Auxologico Italiano, Milan, Italy
[12]Division of Mental Health and Addiction, University of Oslo, Oslo, Norway
[13]University of Nicosia, Nicosia, Cyprus

Corresponding Author:
Shaun Liverpool, MSc
Evidence-Based Practice Unit
University College London and Anna Freud National Centre for Children and Families
4-8 Rodney Street
London, N1 9JH
United Kingdom
Phone: 44 020 3108 9888
Email: shaun.liverpool.14@ucl.ac.uk

Abstract

Background: There is a high prevalence of children and young people (CYP) experiencing mental health (MH) problems. Owing to accessibility, affordability, and scalability, an increasing number of digital health interventions (DHIs) have been developed and incorporated into MH treatment. Studies have shown the potential of DHIs to improve MH outcomes. However, the modes of delivery used to engage CYP in digital MH interventions may differ, with implications for the extent to which findings pertain to the level of engagement with the DHI. Knowledge of the various modalities could aid in the development of interventions that are acceptable and feasible.

Objective: This review aimed to (1) identify modes of delivery used in CYP digital MH interventions, (2) explore influencing factors to usage and implementation, and (3) investigate ways in which the interventions have been evaluated and whether CYP engage in DHIs.

Methods: A literature search was performed in the Cochrane Library, Excerpta Medica dataBASE (EMBASE), Medical Literature Analysis and Retrieval System Online (MEDLINE), and PsycINFO databases using 3 key concepts “child and adolescent mental health,” “digital intervention,” and “engagement.” Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were followed using rigorous inclusion criteria and screening by at least two reviewers. The selected articles were assessed for quality using the mixed methods appraisal tool, and data were extracted to address the review aims. Data aggregation and synthesis were conducted and presented as descriptive numerical summaries and a narrative synthesis, respectively.

Results: This study identified 6 modes of delivery from 83 articles and 71 interventions for engaging CYP: (1) websites, (2) games and computer-assisted programs, (3) apps, (4) robots and digital devices, (5) virtual reality, and (6) mobile text messaging. Overall, 2 themes emerged highlighting intervention-specific and person-specific barriers and facilitators to CYP's engagement. These themes encompassed factors such as suitability, usability, and acceptability of the DHIs and motivation, capability, and opportunity for the CYP using DHIs. The literature highlighted that CYP prefer DHIs with features such as videos, limited text, ability to personalize, ability to connect with others, and options to receive text message reminders. The findings of this review suggest a high average retention rate of 79% in studies involving various DHIs.

Conclusions: The development of DHIs is increasing and may be of interest to CYP, particularly in the area of MH treatment. With continuous technological advancements, it is important to know which modalities may increase engagement and help CYP who are facing MH problems. This review identified the existing modalities and highlighted the influencing factors from the perspective of CYP. This knowledge provides information that can be used to design and evaluate new interventions and offers important theoretical insights into how and why CYP engage in DHIs.

(J Med Internet Res 2020;22(6):e16317) doi:10.2196/16317

KEYWORDS

mHealth; eHealth; technology; smartphone; children; adolescent mental health; mobile phone

Introduction

Prevalence of Mental Health Problems in Children and Young People

Mental health (MH) problems in childhood and adolescence are of great importance because of their prevalence, early onset, and impact on different areas of the child's life [1]. The number of children and young people (CYP) who experience MH problems ranges from 10% to 20% worldwide [2]. An international study conducted in 27 countries estimated the worldwide pooled prevalence of MH problems to be 13.4% among CYP [3]. Specifically, anxiety and disruptive behavior disorders seem to be the most frequent presentations (Table 1) [3]. Estimates further suggest that approximately 1 in every 3 adolescents will meet the criteria for anxiety and depressive disorders [2], while 1 in 4 young people aged 16 to 24 years has experienced at least one MH problem in the past year [3].

Table 1. Prevalence of mental health problems in children and young people.

Mental health problem	Prevalence (%) (adapted from Polanczyk et al [3])
Anxiety	6.5
Disruptive behavior	5.7
Oppositional defiance disorder	3.6
Attention-deficit hyperactivity	3.4
Depression	2.6
Conduct	2.1

Benefits of Digital Health Interventions

Addressing MH problems in CYP is a major public health concern [4,5], which has been impaired by low levels of youth help-seeking behavior [6]. Concerns about stigma and confidentiality, shame or embarrassment in discussing personal issues, financial costs, and/or limited access to services are among the many barriers to accessing help in this population [6-8]. In many instances, existing efficacious face-to-face interventions are adapted using digital technology as a means of addressing these barriers [9]. Digital health interventions (DHIs; eg, internet programs, apps, virtual reality environments, robotic systems) have the potential to be effective, with advantages of accessibility, anonymity, prompt feedback, cost-effectiveness, applicability in real-life contexts, and high treatment fidelity [7,10-15]. Therefore, considering the increased digital literacy and internet use among youth [16], DHIs may serve as a new way to increase accessibility to MH interventions in this population [17,18].

Efficacy of Digital Health Interventions

The World Health Organization (WHO), the United Kingdom's National Health Service, and the US National Institute of Mental Health have identified MH apps as cost-effective and scalable solutions for addressing the MH treatment gap [19]. The efficacy of web-based therapies is well established in the treatment of several MH problems, including depression, anxiety, and substance misuse among adolescents [20-22]. Web-based treatment programs have also demonstrated comparable efficacy to face-to-face psychotherapy [23,24]. In addition, smartphone-based MH interventions have been shown to be a promising self-management tool for depression [25] by reducing symptoms, similar to face-to-face interventions. Equivalent results were also found for anxiety-focused mobile apps [26]. Recent systematic reviews have shown that interventions based on computerized cognitive behavioral therapy are a promising and acceptable way to reduce anxiety and depression in CYP [17,18]. The findings also support the clinical benefits of DHIs for other symptoms and disorders such as autism spectrum disorders, attention deficit, and behavioral disorders [27].

Research on Digital Mental Health Interventions for Children and Young People

Despite the growing interest in using mobile apps to deliver interventions, more research evidence is needed to support implementation in children and young people’s mental health (CYPMH) services [27-29]. For instance, the evaluation of DHIs is increasingly discussed in electronic health research. A recent review showed that the majority of registered DHI evaluation trials employ common methods, such as the randomized controlled trial (RCT) study design [30]. There is much debate in the literature on appropriate methods for evaluating the impact of DHIs [31]. For instance, given the speed at which technologies advance, adaptive research designs may be more useful for increasing usability and ability to respond in a timely manner to users’ experiences [32]. Considering the limitations of traditional research designs, new methodological frameworks and research designs have been developed (eg, continuous evaluation of evolving behavioral intervention technologies [33] and microrandomized trials [34]). To develop DHIs that are more useful and thereby more engaging for users, researchers agree that the impact of different functionalities on levels of engagement is important [35]. Engagement with digital behavior change interventions has been defined in the literature as a subjective experience (the user-perceived state of *flow*, characterized by temporal dissociation, focused attention, interest, and enjoyment) or as a behavior (extent of usage of the DHI over time or adherence) [36]. Perski et al [36] proposed an integrated conceptualization of engagement that includes both the extent (eg, amount, frequency, duration, and depth) of usage and the subjective experience of “what it feels” to be engaged (eg, attention, interest, and affect). However, engagement is usually assessed through the evaluation of the user interaction with the DHI, either by user-reported tools (eg, questionnaires, interviews, or think-aloud studies), by automated recordings of use (eg, log-ins, page views), or by recording physiological or psychophysical correlates of DHI interaction [36].

Despite the potential of DHIs, researchers have identified several limitations that influence practicality [37]. The main limitations identified were restricted tailoring to patient needs, challenges with managing comorbidity and acute crisis [38], low patient engagement and high dropout rates [39]. Although efforts have been made to reduce these occurrences with strategies such as gamification, tailoring, and guided self-help, the aforementioned difficulties remain [27,38,39]. In addition, further challenges arise from cautious attitudes of professionals toward DHIs, such as failure to address important aspects of the disease, data security, and accessibility [40].

Rationale for This Review

The rapid advancement of technology [41] and the increasing interest of CYP in technology [16] calls for a better understanding and evaluation of DHIs used to *engage* CYP with MH problems [42,43]. Engagement is commonly referred to as the active involvement of participants with the intervention, also described in previous literature as *participation*, *adherence*, *noncompliance*, or *resistance* [44]. This knowledge is crucial to support the development and evaluation of DHIs that are acceptable and feasible in CYPMH settings. This review sets out to contribute to the growing body of knowledge on digital CYPMH interventions by investigating modes of delivery used in DHIs. Although recent meta-analyses highlighted the potential effectiveness of CYPMH DHIs [17,18,45], this review aimed to present information that might be of use in the development of real-world interventions that are more likely to increase engagement from CYP.

Aims

The primary aim of this study was to review the literature to identify modes of delivery used to engage CYP in digital MH interventions. Second, we explored barriers and facilitators for the usage and implementation of DHIs. The authors also aimed to investigate the ways in which these interventions have been evaluated and whether CYP engage in DHI research. The following questions were addressed:

1. What modes of delivery are used for engaging CYP in digital MH interventions?
2. What are the barriers and facilitators to engaging CYP in digital MH interventions?
3. How do retention rates vary in CYP digital MH intervention research?

Methods

Literature Search and Search Strategy

A literature search was conducted using the Cochrane Library, EMBASE, MEDLINE, and PsycINFO databases. All searches were carried out on the same day (December 27, 2018) to control for daily updates. Overall, 3 key concepts informed the search strategy: *child and adolescent mental health*, *digital intervention*, and *engagement*. Terms within similar categories were combined with *OR* and then the results from each category were combined with *AND* (see Multimedia Appendix 1). The search strategy was guided by similar reviews exploring technology or engagement in child and family MH treatment [44-46], the review team discussions, and input from the University College London Institute of Child Health librarian. Reference lists of relevant articles were also scanned for additional potential studies. An initial sample of articles identified through database searching was screened first by titles and abstracts. Next, the full-text versions of potentially relevant studies were retrieved and examined in detail for eligibility at the review team meetings. Differences regarding study selection were resolved by discussion among the authors.

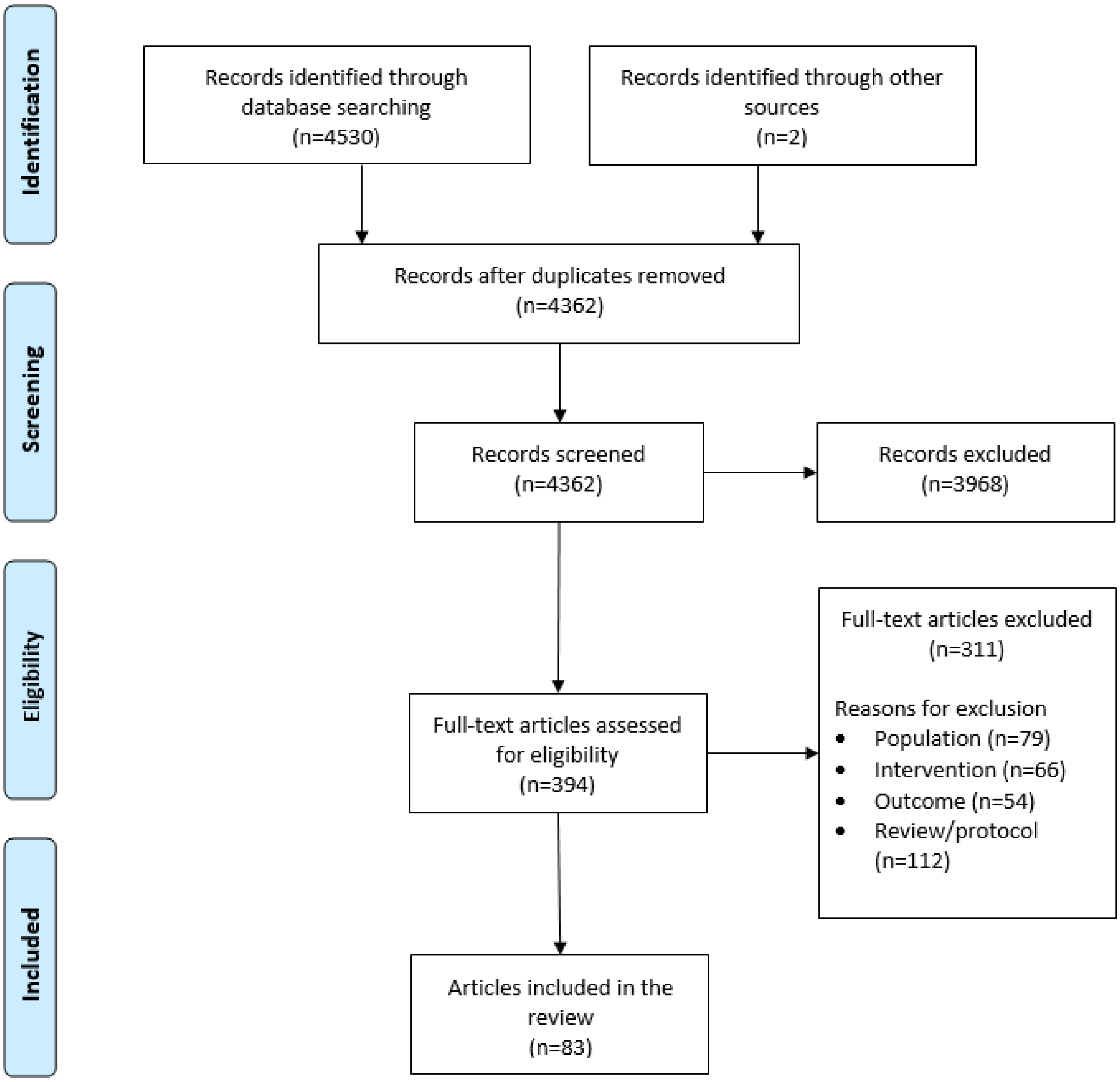
Inclusion and Exclusion Criteria

Screened articles were included if (1) the study targeted a CYP sample with a mean age less than 25 years; (2) the article described a DHI targeting an MH symptom (related to a primary physical/somatic condition) or the intervention was being used by CYP with MH problems; and (3) the study explored the development or testing of a DHI resulting in data on adherence, acceptability, or barriers and facilitators to engagement. Any study design was deemed acceptable for inclusion. Articles were excluded if (1) the age of the participants was not defined or if the mean age of the sample was 25 years and above, (2) the intervention was for the sole purpose of communicating between a health care professional and the CYP (eg, Skype, email, teleconference, or messages for appointment reminders), (3) the outcome of the study was not clearly defined or did not provide sufficient details to determine if the outcome was directly related to the intervention, and (4) the study had no human participants (eg, discussion articles describing a novel intervention).

Study Selection Process

In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [47], the flowchart presented in Figure 1 provides step-by-step details of our study selection process. After duplicates were removed, at least two members of the review team independently screened the titles and abstracts against the inclusion criteria. The full-text versions of the remaining potential articles were further examined by at least two reviewers for final inclusion. Articles excluded at this stage described interventions that were being used for communication purposes only, did not provide sufficient details of the intervention, were reviews or study protocols, were those in which the age range was not specified or was above the cutoff, targeted a non-MH condition only or targeted parents or clinicians, or had an outcome that was not related to the intervention. Any disagreements were resolved through discussions.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart of the study selection process.



Data Extraction and Quality Assessment

A standardized form [48] that was adapted and piloted by the review team was used to extract relevant information from each article, including the following: reference, year, country, study aims, study design, sample size, setting, clinical characteristics, type of support including therapeutic treatment, retention rate, outcome, and descriptive characteristics of the sample and the intervention. Study-specific data for the second review question were also extracted at this stage to inform the thematic framework [49]. The mixed methods appraisal tool (MMAT-v2018) [50] was used to assess the methodological quality of each selected study. This tool was discussed in detail and selected based on its ability to report on the quality of varying study designs. Responses were rated on a categorical scale as "no," "can't tell," or "yes" to any of the methodological quality criteria. The number of items rated "yes" was counted to provide an overall score out of a possible 5 [51]. If at least one of the MMAT quality criteria was met, the methodological quality of the study was considered acceptable and the record was included. SL and a second member of the review team independently extracted all data and independently conducted the quality assessment. The 2 reviewers discussed any discrepancies and, if necessary, consulted a third team member to reach a final decision.

Data Aggregation and Synthesis

The extracted data were collated and summarized to produce a narrative summary of the study characteristics that addressed the first review question. A descriptive numerical summary was presented to group articles by the primary digital platform used to deliver the intervention. SL completed a qualitative data-driven thematic analysis [52] in addition to inductive analysis informed by the Digital Behaviour Change Framework [53] to address the second review question. Moreover, this framework and the capability, opportunity, motivation, and behavior (COM-B) model [54] were used to explore the factors influencing behavior change and intervention design. The coding process involved moving backward and forward between the data and emerging concepts. The first step generated initial codes from open coding, in which units of meanings were derived from a line-by-line analysis followed by axial coding to integrate and differentiate among subcategories. Qualitative findings relating to barriers and facilitators were coded in NVivo [55]. The review team reviewed the coding process, and any disagreements were discussed before reaching a consensus. There were no major disagreements, and consequently, the codes were developed into themes. For the purpose of addressing the third review question, the retention rate was defined as the percentage of participants completing outcome measures for at least one follow-up time point. In studies where this was not explicitly mentioned, we used the percentage of participants continuing to engage with the intervention after a specified period (ie, a time period identified by the original author).

Changes to the Protocol

Initially, the review team planned to investigate recruitment rates. However, the identified studies varied in recruitment strategies and did not provide sufficient details to address this research question. In addition, although the review team acknowledged the potential of gray literature (eg, research not published in peer-reviewed journals) to broaden the scope of systematic reviews, the team agreed to only include articles published in peer-reviewed journals. This decision was because of various reasons: (1) the popularity of technological advancements in health care; (2) the resource constraints of this study; (3) some evidence of the limited contribution of unpublished studies to the results of meta-analyses in child-relevant reviews [56]; and (4) the consideration that the aim of this systematic review was not related to efficacy and safety, which could be amenable to publication bias. No other substantial deviations from the registered protocol were made. The review protocol was registered in the International Register of Systematic Reviews (PROSPERO) [CRD42018094815].

Results

Overview of the Included Articles

The results of this systematic review are presented as a narrative synthesis [57] and, where applicable, descriptive numerical summaries are provided. A total of 83 articles published between 2001 and 2018 met the inclusion criteria (Multimedia Appendix 2) identifying 71 interventions (Multimedia Appendix 3). Of the 83 articles reviewed, almost two-third were conducted in the United States and Canada or in Australia and New Zealand. The most common type of intervention incorporated cognitive behavioral therapy as the main therapeutic modality. The mean age of the included samples was between 2 years to 24 years. Affective disorders (ie, anxiety and depression, including suicidality) were the most common presentation targets of the DHIs reviewed. Multimedia Appendix 4 provides details of the reviewed articles and Table 2 provides a summary of these findings.

A broad range of recruitment strategies were used to develop and test these DHIs, including referrals from health or school professionals; self-referrals through social media and web-based advertising; university email lists; recruitment software; or in-person advertising through posters, flyers, newspaper advertisements, word of mouth, and existing research and support groups. The following section presents the modes of delivery for DHIs, highlighting how they have been evaluated and their main features and purpose. Table 3 provides a summary of the 6 DHI categories identified in this review, the corresponding features, and the study designs adopted.

Table 2. Characteristics of the included articles (N=83).

Characteristics	Values, n (%)
Country	
United States and Canada	31 (37)
Australia and New Zealand	23 (28)
Europe	21 (25)
Asia	7 (8)
Brazil	1 (1)
Therapeutic modality	
Cognitive behavioral therapy	39 (47)
Cognitive skills training mechanisms	9 (11)
Social skills training or social support	7 (8)
Applied behavior analysis concepts	3 (4)
Single component or combinations[a]	25 (30)
Disorders	
Affective[b]	38 (46)
Attention deficit and hyperactivity	7 (8)
Autism spectrum	12 (15)
Eating disorders	4 (5)
Behavioral disorders[c]	10 (12)
Nonspecific or multiple disorders	12 (15)

[a] Combinations of the following strategies: therapeutic support embedded in positive psychology, behavior activation, self-regulation, learning theories, motivational interviewing, and mindfulness.

[b] Depression, anxiety, or suicidality.

[c] Obsessive compulsive disorder, substance abuse, selective mutism, social difficulties, and psychosis or schizophrenia.

Table 3. Summary of digital modes of delivery used in children and young people's mental health intervention.

Mode of delivery (number of articles)	Goals: features	Study design, n
Website interventions (n=43)	• Communication: emails, text messages, social networking, web-based message boards, discussion forums • Dissemination of information: text and multimedia channels (videos, animations, and audio), games and quizzes, homework tasks, and web-based profile set up with customizations	RCT[a] (n=22), pre- to posttest (n=11), observational study—qualitative, quantitative, or mixed methods approaches (n=10)
Games or computer-assisted interventions (n=23)	• Dissemination of information, skills development, psychoeducation: photos, stories, animations, quizzes, text messages, and multimedia (audio and videos)	RCT (n=11), pre- to posttest (n=8), observational study—quantitative or mixed methods approaches (n=4)
Apps: web or mobile (n=10)	• Dissemination of information, skills development, peer-to-peer communication: text message reminders, text, photos and multimedia (audio and videos) plus an opportunity to upload content	RCT (n=4), pre- to posttest (n=1), observational study—qualitative, quantitative, or mixed methods approaches (n=5)
Robots and digital devices (n=3)	• Dissemination of information: audio and movement • Peer-to-peer communication: email reminders	RCT (n=1), feasibility study (n=1), mixed methods design (n=1)
Virtual reality experiences (n=3)	• Dissemination of information, skills development, therapeutic support: Gamification using multimedia (audio and images)	Pre- to posttest (n=2), posttest (n=1)
Mobile text messages (n=1)	• Dissemination of information, skills development, social support: Text	RCT (n=1)

[a] RCT: randomized controlled trial.

What Modes of Delivery Are Used for Engaging Children and Young People in Digital Mental Health Interventions?

Website Interventions

Overall, 33 of the 43 articles adopted an interventional study design (22 RCTs and 11 pre- to poststudy designs). The remaining 10 studies adopted observational study designs utilizing qualitative, quantitative, or mixed methods approaches. The methodological quality of the included studies was acceptable. Qualitative studies scores ranged from 2 to 5 points, RCTs and nonrandomized quantitative studies also ranged 2 to 5 points, and mixed methods studies ranged 2 to 4 points.

The primary goal of the majority of the interventions (n=40) was to transmit specific MH information to a targeted population. In addition, 12 of the 40 articles described interventions that were multipurpose, providing an additional opportunity for peer communication (n=7) or for personal health tracking (n=4). YouthCHAT (see Multimedia Appendix 3 for definitions and descriptions of the interventions) provided general information in addition to providing an opportunity for personal health tracking. However, SharpTalk's primary focus was to facilitate peer-to-peer communication through discussion forums, and Manage Your Life Online functioned as a communication aid that provided an opportunity for personal health tracking.

Various features were adopted to achieve the above goals. Communication occurred digitally using emails, text messages, social networking, web-based message boards, and discussion forums. Dissemination of information occurred through text and multimedia channels (eg, videos, animations, and audio). Some interventions also utilized games and quizzes, homework tasks, and a web-based profile set up with customizations.

Games or Computer-Assisted Interventions

Overall, 20 of the 23 articles adopted an interventional study design (11 RCTs and 8 pre- to poststudy designs). The remaining 4 studies adopted observational study designs utilizing quantitative or mixed method approaches. The methodological quality of the included studies varied. RCTs scores ranged 2 to 5 points, nonrandomized quantitative studies ranged 2 to 4 points, and mixed methods ranged 3 to 5 points. No articles used qualitative methods only.

The primary goal of the majority of the interventions (n=18) was to transmit specific MH information to a targeted population. Of the 18 interventions, 4 were multipurpose, providing additional general information to the public (n=1) or an opportunity for personal health tracking (n=4). In addition, 8 interventions focused on cognitive training tasks. The computer-assisted instruction intervention was used as a facilitator to assist children in developing reading skills. The social stories accessed via tablets were also used for psychoeducational purposes.

The gamification approach used to achieve the above goals was accessed either on the web or offline and incorporated photos, stories, animations, quizzes, text messages, and videos.

Apps

Of the 10 articles, 5 adopted an interventional study design (4 RCTs and 1 pre- to poststudy design). The remaining 5 articles adopted observational study designs utilizing qualitative, quantitative, or mixed methods approaches. The methodological quality of the included studies varied. Qualitative studies scored either 4 or 5 points, RCTs scores ranged 1 to 4 points, the 1 nonrandomized quantitative study scored 4 points, whereas the 2 mixed methods studies scored 3 points.

The primary goal of most apps was to transmit specific MH information to a targeted population. Furthermore, 3 apps were multipurpose, providing an additional opportunity for personal health tracking. The TECH app further included peer-to-peer communication.

The included apps were either web apps or mobile apps and included text message reminders, text, photos, and multimedia (videos and audio). Users also had the opportunity to upload content such as videos and photos.

Robots and Digital Devices

Of the 3 studies, 2 adopted an interventional study design, of which one (CommU) was an RCT. The study on the Fitbit Flex and Facebook adopted a mixed methods design. The study on ARIA adopted an observational study design as a pilot usability study. The methodological quality of the included studies was acceptable. The ARIA study scored 3 points, CommU scored 3 points, and the Fitbit Flex and Facebook intervention study scored 5 points.

The primary goal of ARIA and CommU was to transmit specific MH information to a targeted population, whereas the Fitbit Flex and Facebook intervention additionally provided an opportunity for peer communication.

ARIA and CommU utilized audio and movement to achieve the above purpose. The Fitbit Flex and Facebook synced with 2 other approaches, an app and a website, and included email reminders to achieve its purpose.

Virtual Reality Experiences

Furthermore, of the 3 studies, 2 interventions adopted pre- to posttest designs, whereas one adopted a posttest-only design (Virtual Dolphin Interaction). No RCTs were found to evaluate the identified interventions. Cave Automatic Virtual Environment also incorporated a mixed methods approach and obtained qualitative data. The methodological quality was acceptable. The Cave Automatic Virtual Environment and Collaborative Virtual Environment studies scored 3 points and the Virtual Dolphin Interaction study scored 4 points.

The primary goal of all 3 interventions was to transmit specific MH information to a targeted group or to facilitate skills training or provide therapeutic support. Collaborative Virtual Environment also functioned as a communication aid to facilitate collaboration within the virtual reality environment.

Cave Automatic Virtual Environment, Collaborative Virtual Environment, and Virtual Dolphin Interaction utilized features of the gamification approach to engage CYP in a more real-life experience, allowing for more immersion and movement.

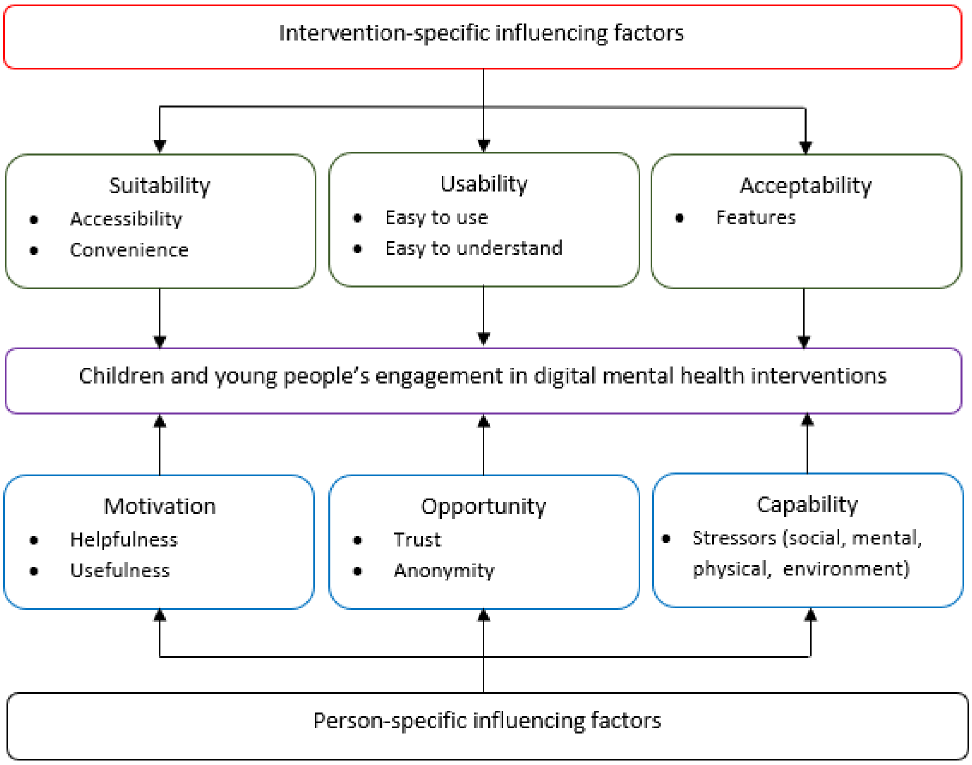
Mobile Text Messages

In addition, 1 text messaging intervention was identified and evaluated in an RCT. The methodological quality score was 3. The Educating and Supporting Inquisitive Youth in Recovery program aimed to transmit specific MH information to a targeted audience, to facilitate skills training or offer therapeutic support, to provide the opportunity for personal health tracking, and to signpost CYP to additional social support websites. No additional features, apart from text, were described in the study. However, participants were contacted via phone as part of the study.

What Are the Barriers and Facilitators to Engaging Children and Young People in Digital Mental Health Interventions?

Influencing factors presented as barriers and facilitators to engagement emerged as 2 broad themes encompassing 6 factors: intervention-specific influences (suitability, usability, and acceptability) and person-specific influences (motivation, capability, and opportunity). Overall, 29% (24/83) of the included articles provided data to support these themes, 13 studies provided data for suitability, 13 for usability, and 14 provided data for acceptability. Of the 24 articles, 8 provided data to inform motivation, 4 for capability, and 13 for opportunity. A summary of concepts corresponding to the individual factors within the major themes is presented in Multimedia Appendix 5. On the basis of these findings, a conceptual framework (Figure 2) was developed, highlighting the specific components impacting engagement in CYP digital MH interventions, which can inform the development of, and research into, CYP DHIs.

Figure 2. A framework of factors influencing engagement in children and young people's mental health digital intervention.



Theme 1: Intervention-Specific Influences

CYP highlighted factors related to the development of the intervention, which influenced whether they used the intervention or not. A prominent factor influencing the acceptability (ie, willingness to use) of the intervention was the features, whereby CYP highlighted certain images, specific language, and interfaces that were unappealing to them. They made suggestions and highlighted features such as videos, having less text, ability to personalize or create a profile, and ability to connect with others or receive text message reminders as encouraging their use of the intervention. CYP also suggested that providing rewards could also be a motivating factor for engaging with DHIs. Similarly, usability (ie, the degree to which the DHI was able to be used) was important for promoting engagement. Interventions that CYP favored were described as self-paced, user friendly, age appropriate, simple, and straightforward. However, in situations where CYP had problems understanding the task, or if the intervention did not provide sufficient instructions on usage, they were less inclined to continue using the DHI.

Another main factor was the suitability (ie, the degree to which the DHI is in line with daily activities) of the intervention to the lifestyle of CYP. Although CYP liked not having to travel to access the intervention and the ability to use it while at home, they were *put off* by technical issues or having to use media such as emails or desktop computers that they used less frequently in their daily lives. Many CYP highlighted not participating in the intervention because of a lack of time and inability to integrate the task into their everyday life. They suggested that flexibility concerning time and ability to bypass long waiting lists encouraged usage. In addition, they highlighted that DHIs were convenient and welcomed as they spend most of their time on the web.

Theme 2: Person-Specific Influences

Of the 6 factors, 3 were associated with person-specific barriers and facilitators to behavior change, which is in line with the COM-B model [54]. The opportunity for the intervention to be adopted was highlighted in 3 areas. First, feeling a sense of connectedness was important to CYP. They were more likely to use the intervention if it facilitated conversations with others because they wanted to know that others had similar experiences. Some CYP even indicated that they “felt alone” on the web without the support of a therapist. Trust was also of great importance to CYP, and they were reluctant to accept DHIs because of privacy concerns or uncertainties around its validity. CYP made suggestions to use trusted *brand names* that they were familiar with. They were also more inclined to use the interventions if there was transparency or evidence provided to support its credibility. The concerns around trust also extended to their preference for anonymity. They highlighted that anonymity made it easier to talk to a stranger on the web without feelings of embarrassment.

The second major factor identified as a person-specific influence was motivation. Some tasks were of less interest to CYP, and sometimes, they would have forgotten the existence of the intervention. However, they highlighted that curiosity and perceived need influenced their usage. Perceived usefulness of the intervention to address their needs was a major motivating factor; therefore, if the resource was viewed as unhelpful or too general, CYP were less interested in using it. However, although CYP were eager to use DHIs, the capability to engage with the intervention was sometimes affected by physical, environmental, and MH stressors, representing another major influencing factor.

How Do Retention Rates Vary in Children and Young People Digital Mental Health Intervention Research?

Owing to the heterogeneity in study design and intervention type, not all articles provided sufficient detail to estimate retention rates. Therefore, we were unable to include estimates for 16 of the included articles. The average retention rate across the remaining 67 articles showed almost 80% of CYP using DHIs or completing the follow-up measures. Results showed that the retention rates varied, with 11.9% (8/67) of articles achieving 100% retention and 8.95% (6/67) reporting less than 50% retention. Overall, 83.58% (56/67) of the included articles had a retention rate of at least 70%. Subgroup analyses for approaches with a larger number of articles indicated that the average retention rate for games and computer-assisted intervention studies was 86.95%, followed by websites interventions with 78.87%, and apps with 78.45%. Multimedia Appendix 4 shows the distribution of retention rates across studies.

Discussion

Overview

This review identified 6 modes of delivery of DHIs for CYP with MH issues: websites, games and computer-assisted programs, apps, robots and digital devices, virtual reality, and mobile text messaging. Overall, 2 themes emerged, highlighting intervention-specific and person-specific barriers and facilitators to engagement in DHIs. In addition, the findings of this review suggest a high average retention rate of almost 80% when the identified modalities were evaluated. Knowledge of these approaches, including influencing factors to usage from the perspective of the CYP, provided information that can be used to design and evaluate new DHIs.

Explanation of the Overall Findings

From the articles reviewed, 59 contributions were published between 2013 and 2018, with 15 records published in 2018. This is consistent with the fact that interest in applying digital technologies to MH practice has been increasing since the early 2000s, and recommendations for research in this field were issued only in 2013 [58,59]. With most of the studies reviewed being conducted in developed countries, digital responses to CYPMH seem unbalanced. Previous research [60-62] highlighted the paucity of access to DHIs in low- and middle-income countries. This inequality could be because of limited resources (both financial and human), shortage of skilled personnel, infrastructure problems leading to poor internet penetration and connectivity [63], or the absence of a specific CYPMH policy [64]. However, a strong association between the severity of risk for mood disorders and social disadvantage has been documented [65,66]. Anxiety is the most common psychiatric condition affecting CYP in all societies [45,67-69]. It may also co-occur with other disorders, both concurrently and sequentially, leading to further health problems [68]. Therefore, it is not surprising, that 46% of articles resulting from our literature search targeted CYP with affective disorders.

The modes of delivery identified in this review are similar to those identified in other reviews exploring adolescent physical health [70,71] and MH [45]. The purpose of the interventions is also in line with WHO’s classification of DHIs [72]. The intervention-specific and person-specific influences on CYP engagement identified in the present review mirror those of previous research on the broader technology acceptance model [73-75], the conceptual framework for engagement in Digital Behavior Change Interventions in adults [36,53], and more recently, the application of social cognitive theory to understand engagement with DHIs for trauma recovery [35]. For these models, predictors included perceived need, engagement self-efficacy, outcome expectations, and symptom severity. The proposed model of CYP engagement in digital MH interventions, based on the findings of the present review, builds on these models by highlighting the importance of the social context in which young people engage with DHIs. It also highlights major factors for researchers and developers in CYPMH to facilitate opportunities for a sense of connectedness. *Peer -connectedness* may be challenged by the necessary application of safeguarding measures. Indirect peer connectedness where CYP can access appropriately anonymous and asynchronous stories from real CYP with similar experiences may be one such approach, as in SharpTalk [76]. Similarly, options for anonymous usage may be particularly important for CYP's *self-connectedness* in terms of sharing experience in a manner that mitigates the role of stigma and shame [77,78]. *Professional connectedness* was particularly important in DHIs as CYP wanted to connect with a trusted support provider in lieu of connecting with a professional in real life. A key facilitator of professional connectedness was credibility in relation to evidence of the intervention's effectiveness and trust in the privacy and data security, which could be facilitated by using familiar *brand names*. Although characterized as an intervention-specific factor in other models [53], we characterized this as part of opportunities for connectedness and, therefore, as a person-specific factor. Future studies should explore the impact of new modes of delivery to promote a sense of connectedness in DHIs (eg, more usage of features such as ChatBots, as in the Manage Your Life Online intervention).

Other barriers and facilitators that were identified in this review also emphasize the importance of user-centered design methods when developing DHIs for CYP [53]. Through co-design workshops and focus groups with CYP, developers can ensure that a DHI's design is age appropriate, (eg, little text and using youth-engaging language) by putting a greater focus on videos and pictures, while keeping the platform user friendly. Moreover, CYP mentioned factors such as reward systems and reminders, which fall under the umbrella of persuasive design methods and have been explored in previous research [36,79]. The positive influence of these methods on user engagement and adherence to DHIs has been supported; however, quantifiable evidence from trials is still lacking [36,79]. With respect to reminders, past research has indicated a positive impact on engagement. However, excessive and undue reminders have also been shown to have opposing effects [80]. Evidence from previous studies has suggested that specific behavior change techniques, such as goal setting or self-monitoring tools, relate to higher engagement [81]. This review did not extensively investigate these techniques and therefore cannot fully suggest their potential positive effects on the engagement of CYP. However, the findings of this review justify that designing DHIs with CYP in mind would be ideal to promote usage, adherence, and positive user experience and to address the barriers that some of the reviewed studies suggest.

Comparison of Research Retention Rates With Other Studies

Our findings suggest that the retention rate of CYP in digital MH interventions (mean retention rate of 79.2%) was superior to that reported in face-to-face CYP MH outpatient care, where dropout affects 20% to 60% of the cases [82]. However, a direct comparison with other studies is not clear, given the diversity of criteria used for defining dropout. For instance, dropout percentages are lower when dropout is defined by the therapist's opinion than when dropout is defined by the completion of a certain number of sessions [83-85]. Our definition of retention relied on completion of the first follow-up measure or engagement for a specified period, which may explain the higher retention found. In this review, retention rates also varied widely across the studies (range 15.79%-100%). A similar heterogeneity in retention rates was found in previous reviews of studies with adults receiving internet-based MH programs (17%-98%) [82], as well as in face-to-face MH interventions with CYP [83], and the adult population (varying between 17%-72% and 17%-98%, respectively) [86]. Efficacy studies tend to present lower dropout rates than studies conducted in naturalistic settings [83]. Our review included a variety of empirical studies, which may have contributed to the diversity of retention rates found. Finally, the average retention rate for games and computer-assisted intervention studies was almost 10% higher than that of the other modalities, which may reflect the preference of children for interventions in game formats [87,88].

Digital Mental Health Care and Support of Children and Young People

DHIs were included at various stages of the provision of psychological support. Technology-mediated programs and tools were part of prevention, assessment, treatment (psychoeducation and psychotherapy) and follow-up of MH care. This extensive potential of DHIs can support the WHO's initiative to identify and intervene to lessen the MH treatment gap [89]. When used as part of the initial assessment, support for shared decision making, personalized goal setting, progression, or management of transitions, DHIs are able to support CYP by enhancing their sense of agency and control. This may in turn promote greater involvement in the treatment process [90]. In several of the reviewed studies, DHIs targeting social skills training and joint attention training were used in the initial phase as a facilitator of the therapeutic process. For some specific conditions such as social anxiety, selective mutism, autism spectrum disorders, and attention-deficit conditions, the involvement of digitized programs in the preliminary phase of therapy may be an important facilitator for therapeutic success [91,92]. DHIs as part of the therapeutic process can be used in the periods between face-to-face treatment for interactive homework assignments, reminders, self-monitoring tools, individualized exercises, and real-time symptom assessment.

Implications and Recommendations

DHIs can be a helpful way to support and treat MH problems. Such tools can complement the various stages of the provision of psychological support or psychotherapy among CYP with MH problems. However, effective implementation and sustained usage will rely on the extent to which the design is appropriate for the intended purpose and how it will be used in practice. This understanding may help to minimize the risks associated with fear of usage that some end users experience by providing useful directions on how to design technologically responsible therapeutic approaches [93]. As a result, the findings of this review suggest that the development of DHIs should be suitable for CYP's lifestyle, focusing on ease of access, such as the ability to be used on their mobile devices at their convenience. Attention should also be given to the design of DHIs to ensure that it is not too complex and that the features are attractive to CYP. In addition, incorporating concepts that provide the CYP with a level of trust for the DHI and the ability to connect with others should be carefully considered. To target concerns about usefulness, developers should work with clinicians and CYP to ensure that suitable information is provided through the intervention. This collaborative approach can highlight specific ways to encourage CYP to continue engaging or increasing engagement during stressful periods.

Although this study provides insight that is valuable for the development of new interventions, future research should (1) not only focus on the effectiveness but also investigate engagement, taking into account influencing factors, as an important component of research; (2) arrive at a consensus on defining engagement and how it should be measured; (3) provide adequate reporting of recruitment and retention rates; and (4) compare CYP preferences for various modes of delivery or therapeutic approaches. Finally, this study also acknowledges the implications for practice. The findings suggest CYP interest in DHIs, and therefore, (1) efforts to improve engagement may be beneficial to CYPMH outcomes and (2) families including the CYP and clinicians should work together to identify DHIs that are suitable to the CYP's lifestyle.

Strengths and Limitations

This review adhered to established guidelines for systematic reviews [94] and adopted a comprehensive study design carried out by a team of researchers, allowing each stage of the review to be undertaken by at least two independent reviewers. Most importantly, this review highlighted the range of modes of delivery, factors influencing usage, and the variation in study types and retention of CYP in digital MH intervention studies. Our findings contribute to a broader understanding of the CYP DHI literature. However, this review is not devoid of limitations. The review team attempted to identify and include as many articles as possible; however, unknowingly and unintentionally, some papers may have been missed. This can be partly because of the challenges and inconsistencies when defining the construct of engagement resulting in a wide variety of terms used [44]. In addition, unpublished data were not included in the search strategy, which may have impacted the results of this review. Nevertheless, this approach was also seen as a further strength by ensuring that only peer-reviewed interventions were included. Moreover, the study team attempted to group interventions based on digital platforms to describe each approach. However, there may still be some variation within these groupings that make it difficult to categorize. This review was also limited as only subsamples of the total number of included articles contributed to addressing research questions 2 and 3. Therefore, caution was taken when generalizing the findings and drawing overall conclusions. In addition, because of the variability in study designs, we were prudent when averaging retention rates for this review, as follow-up measures were collected at varying time points across the selected records.

Conclusions

DHIs may be of interest to CYP, particularly in the area of MH treatment. Research on retention rates suggests high engagement of CYP in digital MH interventions that may encourage further development of DHIs in the near future. CYPMH services could benefit from this development as the included studies indicate. However, the results of this review highlighted intervention-specific and person-specific factors that influence CYP usage of digital MH interventions that should be considered. With continuous technological advancements, it is desirable to know which modalities may increase usability and adherence to better support CYP facing MH challenges.

Acknowledgments

This study was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement number 722561; TREATme, a European Network on Individualized Psychotherapy Treatment of Young People with Mental Health Disorders under the European Cooperation in Science and Technology which is supported by the EU Framework Programme Horizon 2020 action CA16102; European COMPETE2020 (POCI-01-0145-FEDER-030980); Estonian Research Council grant PUT1518; and Portuguese National funds FCT—Fundação para a Ciência e a Tecnologia, I P (PTDC/PSI-ESP/30980/2017).

Authors' Contributions

SL and JE-C conceptualized the paper. SL, CPM, CMDS, AC, S Carletto, CH, SS, S Cerón, PMP, GP, BM, and JE-C carried out selection of articles, data extraction, data analysis, and writing of the original draft. SL, CPM, CMDS, AC, S Carletto, CH, SS, S Cerón, PMP, GP, BM, RU, NF, and JE-C were involved in the writing, review, and editing.

Conflicts of Interest

JE-C is an author of one of the reviewed articles.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 131 KB - jmir_v22i6e16317_app1.pdf]

Multimedia Appendix 2
Reviewed articles.
[DOCX File , 23 KB - jmir_v22i6e16317_app2.docx]

Multimedia Appendix 3
Intervention characteristics.
[PDF File (Adobe PDF File), 488 KB - jmir_v22i6e16317_app3.pdf]

Multimedia Appendix 4
Characteristics of reviewed articles.
[PDF File (Adobe PDF File), 675 KB - jmir_v22i6e16317_app4.pdf]

Multimedia Appendix 5
Themes and supporting codes.
[PDF File (Adobe PDF File), 361 KB - jmir_v22i6e16317_app5.pdf]

References

1. Perou R, Bitsko R, Blumberg S, Pastor P, Ghandour R, Gfroerer J, Centers for Disease Control and Prevention (CDC). Mental health surveillance among children-United States, 2005-2011. MMWR Suppl 2013 May 17;62(2):1-35. [Medline: 23677130]
2. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child and adolescent mental health worldwide: evidence for action. Lancet 2011 Oct 22;378(9801):1515-1525. [doi: 10.1016/S0140-6736(11)60827-1] [Medline: 22008427]
3. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry 2015 Mar;56(3):345-365. [doi: 10.1111/jcpp.12381] [Medline: 25649325]
4. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in US adolescents: results from the national comorbidity survey replication--adolescent supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
5. Lawrence D, Johnson S, Boterhoven DH, Ainley J, Zubrick S. Australian Government Department of Health. 2015. The Mental Health of Children and Adolescents: Report on the Second Australian Child and Adolescent Survey Of Mental Health and Wellbeing URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/9DA8CA21306FE6EDCA257E2700016945/$File/child2.pdf [accessed 2020-05-06]
6. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
7. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
8. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
9. Wangelin B, Szafranski D, Gros D. Telehealth technologies in evidence-based psychotherapy. In: Luiselli JK, Fischer AJ, editors. Computer-Assisted and Web-Based Innovations in Psychology, Special Education, and Health. New York, USA: Academic Press; 2016:119-140.
10. Lahiri U, Bekele E, Dohrmann E, Warren Z, Sarkar N. Design of a virtual reality based adaptive response technology for children with autism. IEEE Trans Neural Syst Rehabil Eng 2013 Jan;21(1):55-64 [FREE Full text] [doi: 10.1109/TNSRE.2012.2218618] [Medline: 23033333]
11. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord 2012 Jan;6(1):249-262 [FREE Full text] [doi: 10.1016/j.rasd.2011.05.006] [Medline: 22125579]
12. van Voorhees BW, Mahoney N, Mazo R, Barrera AZ, Siemer CP, Gladstone TR, et al. Internet-based depression prevention over the life course: a call for behavioral vaccines. Psychiatr Clin North Am 2011 Mar;34(1):167-183 [FREE Full text] [doi: 10.1016/j.psc.2010.11.002] [Medline: 21333846]
13. Yuan SN, Ip HH. Using virtual reality to train emotional and social skills in children with autism spectrum disorder. London J Prim Care (Abingdon) 2018;10(4):110-112 [FREE Full text] [doi: 10.1080/17571472.2018.1483000] [Medline: 30083244]
14. Alvarez-Jimenez M, Alcazar-Corcoles M, González-Blanch C, Bendall S, McGorry P, Gleeson J. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
15. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]

16. OfCom. 2018. Children and Parents: Media Use and Attitudes Report URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/134907/Children-and-Parents-Media-Use-and-Attitudes-2018.pdf [accessed 2020-05-06]
17. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
18. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
19. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
20. Donovan CL, March S. Computer-based treatment programs for youth anxiety: a systematic review. Psychopathology Rev 2014 Jul 28;a1(1):130-156. [doi: 10.5127/pr.033613]
21. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
22. Reyes-Portillo JA, Mufson L, Greenhill LL, Gould MS, Fisher PW, Tarlow N, et al. Web-based interventions for youth internalizing problems: a systematic review. J Am Acad Child Adolesc Psychiatry 2014 Dec;53(12):1254-70.e5. [doi: 10.1016/j.jaac.2014.09.005] [Medline: 25457924]
23. Christensen H, Batterham P, Calear A. Online interventions for anxiety disorders. Curr Opin Psychiatry 2014 Jan;27(1):7-13. [doi: 10.1097/YCO.0000000000000019] [Medline: 24257123]
24. Sethi S. Treating youth depression and anxiety: a randomised controlled trial examining the efficacy of computerised versus face-to-face cognitive behaviour therapy. Aust Psychol 2013 Feb 14;48(4):249-257. [doi: 10.1111/ap.12006]
25. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
26. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
27. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
28. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
29. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
30. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mhealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
31. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
32. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015 Jul 8;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
33. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
34. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
35. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
36. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
37. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based emental health in Europe: the study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]

38. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
39. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014 Nov;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
40. Schuster R, Pokorny R, Berger T, Topooco N, Laireiter A. The advantages and disadvantages of online and blended therapy: survey study amongst licensed psychotherapists in Austria. J Med Internet Res 2018 Dec 18;20(12):e11007 [FREE Full text] [doi: 10.2196/11007] [Medline: 30563817]
41. Ericsson - A World of Communication. 2015. Ericsson Mobility Report: 70 percent of world's population using smartphones by 2020 URL: https://?/www.?ericsson.com/?en/?press-releases/?2015/?6/?ericsson-mobility-report-70-percent-of-worlds-population-using-smartphones-by-2020 [accessed 2020-02-12]
42. Becker KD, Lee BR, Daleiden EL, Lindsey M, Brandt NE, Chorpita BF. The common elements of engagement in children's mental health services: which elements for which outcomes? J Clin Child Adolesc Psychol 2015;44(1):30-43. [doi: 10.1080/15374416.2013.814543] [Medline: 23879436]
43. Gopalan G, Goldstein L, Klingenstein K, Sicher C, Blake C, McKay MM. Engaging families into child mental health treatment: updates and special considerations. J Can Acad Child Adolesc Psychiatry 2010 Aug;19(3):182-196 [FREE Full text] [Medline: 20842273]
44. Haine-Schlagel R, Walsh NE. A review of parent participation engagement in child and family mental health treatment. Clin Child Fam Psychol Rev 2015 Jun;18(2):133-150 [FREE Full text] [doi: 10.1007/s10567-015-0182-x] [Medline: 25726421]
45. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
46. Kim H, Munson MR, McKay MM. Engagement in mental health treatment among adolescents and young adults: a systematic review. Child Adolesc Soc Work J 2012 Mar 3;29(3):241-266. [doi: 10.1007/s10560-012-0256-2]
47. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
48. Munn Z, Aromataris E, Tufanaru C, Stern C, Porritt K, Farrow J, et al. The development of software to support multiple systematic review types: the joanna briggs institute system for the unified management, assessment and review of information (JBI SUMARI). Int J Evid Based Healthc 2019 Mar;17(1):36-43. [doi: 10.1097/XEB.0000000000000152] [Medline: 30239357]
49. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
50. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. Educ Inform J 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
51. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
52. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
53. Kayser L, Nøhr C, Bertelsen P, Botin L, Villumsen S, Showell C, et al. Theory and practice in digital behaviour change: a matrix framework for the co-production of digital services that engage, empower and emancipate marginalised people living with complex and chronic conditions. Informatics 2018 Nov 9;5(4):41. [doi: 10.3390/informatics5040041]
54. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Sutton, UK: Silverback Publishing; 2014.
55. Jackson K, Bazeley P. Qualitative Data Analysis with NVivo. Third Edition. Thousand Oaks, CA: Sage Publications; 2019.
56. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. BMC Med Res Methodol 2017 Apr 19;17(1):64 [FREE Full text] [doi: 10.1186/s12874-017-0347-z] [Medline: 28420349]
57. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. CiteSeerX. 2006. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.3100&rep=rep1&type=pdf [accessed 2020-05-14]
58. Hatch A, Hoffman JE, Ross R, Docherty JP. Expert consensus survey on digital health tools for patients with serious mental illness: optimizing for user characteristics and user support. JMIR Ment Health 2018 Jun 12;5(2):e46 [FREE Full text] [doi: 10.2196/mental.9777] [Medline: 29895514]

59. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
60. Sobowale K, Nguyen M, Weiss B, Van TT, Trung LT. Acceptability of internet interventions for youth mental health in Vietnam. Glob Ment Health (Camb) 2016;3:e22 [FREE Full text] [doi: 10.1017/gmh.2016.18] [Medline: 28596890]
61. Cummings JR, Gaydos LM, Mensa-Kwao A, Song M, Blake SC. Perspectives on caregiver-focused mhealth technologies to improve mental health treatment for low-income youth with ADHD. J Technol Behav Sci 2019 Mar;4(1):6-16 [FREE Full text] [doi: 10.1007/s41347-018-0066-2] [Medline: 31106258]
62. Eustache E, Gerbasi ME, Fawzi MC, Fils-Aimé JR, Severe J, Raviola GJ, et al. High burden of mental illness and low utilization of care among school-going youth in Central Haiti: a window into the youth mental health treatment gap in a low-income country. Int J Soc Psychiatry 2017 May;63(3):261-274 [FREE Full text] [doi: 10.1177/0020764017700174] [Medline: 28367719]
63. Chakrabarti S, Shah R. Telepsychiatry in the developing world: whither promised joy? Indian J Soc Psychiatry 2016;32(3):273. [doi: 10.4103/0971-9962.193200]
64. Shatkin JP, Belfer ML. The global absence of child and adolescent mental health policy. Child Adolesc Ment Health 2004 Sep;9(3):104-108. [doi: 10.1111/j.1475-3588.2004.00090.x]
65. McLaughlin KA, Costello EJ, Leblanc W, Sampson NA, Kessler RC. Socioeconomic status and adolescent mental disorders. Am J Public Health 2012 Sep;102(9):1742-1750 [FREE Full text] [doi: 10.2105/AJPH.2011.300477] [Medline: 22873479]
66. Hartwell H. Social inequality and mental health. J R Soc Promot Health 2008 May;128(3):98. [doi: 10.1177/14664240081280030701] [Medline: 18595622]
67. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
68. Langhaug LF, Pascoe SJ, Mavhu W, Woelk G, Sherr L, Hayes RJ, et al. High prevalence of affective disorders among adolescents living in rural Zimbabwe. J Community Health 2010 Aug;35(4):355-364 [FREE Full text] [doi: 10.1007/s10900-010-9261-6] [Medline: 20571897]
69. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues Clin Neurosci 2009;11(1):7-20 [FREE Full text] [Medline: 19432384]
70. Devine KA, Viola AS, Coups EJ, Wu YP. Digital health interventions for adolescent and young adult cancer survivors. JCO Clin Cancer Inform 2018 Dec;2:1-15 [FREE Full text] [doi: 10.1200/CCI.17.00138] [Medline: 30652583]
71. Rose T, Barker M, Jacob CM, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
72. World Health Organization. 2018. Classification of Digital Health Interventions V1.0: A Shared Language to Describe the Uses of Digital Technology for Health URL: https://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/ [accessed 2020-05-14]
73. Venkatesh V, Morris MG, Davis GD, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
74. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
75. Rahimi B, Nadri H, Afshar HL, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
76. Jones R, Sharkey S, Ford T, Emmens T, Hewis E, Smithson J, et al. Online discussion forums for young people who self-harm: user views. Psychiatrist 2018 Jan 2;35(10):364-368. [doi: 10.1192/pb.bp.110.033449]
77. Kaushik A, Kostaki E, Kyriakopoulos M. The stigma of mental illness in children and adolescents: a systematic review. Psychiatry Res 2016 Sep 30;243:469-494. [doi: 10.1016/j.psychres.2016.04.042] [Medline: 27517643]
78. Stangl AL, Earnshaw VA, Logie CH, van Brakel W, Simbayi LC, Barré I, et al. The health stigma and discrimination framework: a global, crosscutting framework to inform research, intervention development, and policy on health-related stigmas. BMC Med 2019 Feb 15;17(1):31 [FREE Full text] [doi: 10.1186/s12916-019-1271-3] [Medline: 30764826]
79. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
80. Lin H, Wu X. Intervention strategies for improving patient adherence to follow-up in the era of mobile information technology: a systematic review and meta-analysis. PLoS One 2014;9(8):e104266 [FREE Full text] [doi: 10.1371/journal.pone.0104266] [Medline: 25100267]
81. Epton T, Currie S, Armitage CJ. Unique effects of setting goals on behavior change: systematic review and meta-analysis. J Consult Clin Psychol 2017 Dec;85(12):1182-1198. [doi: 10.1037/ccp0000260] [Medline: 29189034]
82. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]

83. de Haan AM, Boon AE, de Jong JT, Hoeve M, Vermeiren RR. A meta-analytic review on treatment dropout in child and adolescent outpatient mental health care. Clin Psychol Rev 2013 Jul;33(5):698-711. [doi: 10.1016/j.cpr.2013.04.005] [Medline: 23742782]
84. Warnick EM, Gonzalez A, Weersing VR, Scahill L, Woolston J. Defining dropout from youth psychotherapy: how definitions shape the prevalence and predictors of attrition. Child Adolesc Ment Health 2011 Jun 15;17(2):76-85. [doi: 10.1111/j.1475-3588.2011.00606.x]
85. O'Keeffe S, Martin P, Goodyer IM, Wilkinson P, Consortium I, Midgley N. Predicting dropout in adolescents receiving therapy for depression. Psychother Res 2018 Sep;28(5):708-721. [doi: 10.1080/10503307.2017.1393576] [Medline: 29084488]
86. Swift JK, Greenberg RP. Premature discontinuation in adult psychotherapy: a meta-analysis. J Consult Clin Psychol 2012 Aug;80(4):547-559. [doi: 10.1037/a0028226] [Medline: 22506792]
87. Heidi P, Minna A, Lotta K, Kai K, Camilla L, Ville L, et al. The advantages and limitations of digital games in children's health promotion. Finnish J Ehealth Ewelfare 2014;6(4):- [FREE Full text]
88. Olson CK. Children's motivations for video game play in the context of normal development. Rev Gen Psychol 2010 Jun;14(2):180-187. [doi: 10.1037/a0018984]
89. World Health Organization. mhGAP : Mental Health Gap Action Programme: Scaling Up Care for Mental, Neurological and Substance Use Disorders. Geneva, Switzerland: World Health Organization; 2008.
90. Edbrooke-Childs J, Edridge C, Averill P, Delane L, Hollis C, Craven MP, et al. A feasibility trial of power up: smartphone app to support patient activation and shared decision making for mental health in young people. JMIR Mhealth Uhealth 2019 Jun 4;7(6):e11677 [FREE Full text] [doi: 10.2196/11677] [Medline: 31165709]
91. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
92. Pihlaja S, Stenberg J, Joutsenniemi K, Mehik H, Ritola V, Joffe G. Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders - a systematic review. Internet Interv 2018 Mar;11:1-10 [FREE Full text] [doi: 10.1016/j.invent.2017.11.005] [Medline: 30135754]
93. George MJ, Odgers CL. Seven fears and the science of how mobile technologies may be influencing adolescents in the digital age. Perspect Psychol Sci 2015 Nov;10(6):832-851 [FREE Full text] [doi: 10.1177/1745691615596788] [Medline: 26581738]
94. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. J R Soc Med 2003 Mar;96(3):118-121 [FREE Full text] [doi: 10.1258/jrsm.96.3.118] [Medline: 12612111]

Abbreviations

COM-B: capability, opportunity, motivation, and behavior
CYP: children and young people
CYPMH: children and young people’s mental health
DHI: digital health intervention
MH: mental health
MMAT: mixed methods appraisal tool
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 11.10.19; peer-reviewed by EB Davies, S Rush, V Ta, C Yeager; comments to author 02.12.19; revised version received 15.01.20; accepted 12.03.20; published 23.06.20.

Please cite as:
Liverpool S, Mota CP, Sales CMD, Čuš A, Carletto S, Hancheva C, Sousa S, Cerón SC, Moreno-Peral P, Pietrabissa G, Moltrecht B, Ulberg R, Ferreira N, Edbrooke-Childs J
Engaging Children and Young People in Digital Mental Health Interventions: Systematic Review of Modes of Delivery, Facilitators, and Barriers
J Med Internet Res 2020;22(6):e16317
URL: http://www.jmir.org/2020/6/e16317/
doi:10.2196/16317
PMID:32442160

Review

Impact of the Method of Delivering Electronic Health Behavior Change Interventions in Survivors of Cancer on Engagement, Health Behaviors, and Health Outcomes: Systematic Review and Meta-Analysis

Kate Furness[1,2,3], BHSc (Hons), BSc; Mitchell N Sarkies[2,3,4,5], BASc, PhD; Catherine E Huggins[6], BSc, PhD; Daniel Croagh[7,8], MBBS, PhD; Terry P Haines[2,3], BPhys, PhD

[1]Nutrition and Dietetics, Monash Medical Centre, Monash Health, Melbourne, Australia

[2]School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[3]Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[4]School of Public Health, Faculty of Health Sciences, Curtin University, Perth, Australia

[5]Centre for Healthcare Resilience and Implementation Science Australian Institute of Health Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia

[6]Department of Nutrition, Dietetics and Food, School of Clinical Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[7]Upper Gastrointestinal and Hepatobiliary Surgery, Monash Medical Centre, Monash Health, Melbourne, Australia

[8]Department of Surgery, School of Clinical Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

Corresponding Author:
Kate Furness, BHSc (Hons), BSc
Nutrition and Dietetics
Monash Medical Centre
Monash Health
246 Clayton Rd
Clayton, Victoria
Melbourne, 3168
Australia
Phone: 61 95944180
Email: kate.furness@monashhealth.org

Abstract

Background: Increased accessibility to the internet and mobile devices has seen a rapid expansion in electronic health (eHealth) behavior change interventions delivered to patients with cancer and survivors using synchronous, asynchronous, and combined delivery methods. Characterizing effective delivery methods of eHealth interventions is required to enable improved design and implementation of evidence-based health behavior change interventions.

Objective: This study aims to systematically review the literature and synthesize evidence on the success of eHealth behavior change interventions in patients with cancer and survivors delivered by synchronous, asynchronous, or combined methods compared with a control group. Engagement with the intervention, behavior change, and health outcomes, including quality of life, fatigue, depression, and anxiety, were examined.

Methods: A search of Scopus, Ovid MEDLINE, Excerpta Medica dataBASE, Cumulative Index to Nursing and Allied Health Literature Plus, PsycINFO, Cochrane CENTRAL, and PubMed was conducted for studies published between March 2007 and March 2019. We looked for randomized controlled trials (RCTs) examining interventions delivered to adult cancer survivors via eHealth methods with a measure of health behavior change. Random-effects meta-analysis was performed to examine whether the method of eHealth delivery impacted the level of engagement, behavior change, and health outcomes.

Results: A total of 24 RCTs were included predominantly examining dietary and physical activity behavior change interventions. There were 11 studies that used a synchronous approach and 11 studies that used an asynchronous approach, whereas 2 studies used a combined delivery method. Use of eHealth interventions improved exercise behavior (standardized mean difference [SMD]

0.34, 95% CI 0.21-0.48), diet behavior (SMD 0.44, 95% CI 0.18-0.70), fatigue (SMD 0.21, 95% CI −0.08 to 0.50; SMD change 0.22, 95% CI 0.09-0.35), anxiety (SMD 1.21, 95% CI: 0.36-2.07; SMD change 0.15, 95% CI −0.09 to 0.40), depression (SMD 0.15, 95% CI 0.00-0.30), and quality of life (SMD 0.12, 95% CI −0.10 to 0.34; SMD change 0.14, 95% CI 0.04-0.24). The mode of delivery did not influence the amount of dietary and physical activity behavior change observed.

Conclusions: Physical activity and dietary behavior change eHealth interventions delivered to patients with cancer or survivors have a small to moderate impact on behavior change and a small to very small benefit to quality of life, fatigue, depression, and anxiety. There is insufficient evidence to determine whether asynchronous or synchronous delivery modes yield superior results. Three-arm RCTs comparing delivery modes with a control with robust engagement reporting are required to determine the most successful delivery method for promoting behavior change and ultimately favorable health outcomes.

(J Med Internet Res 2020;22(6):e16112) doi:10.2196/16112

KEYWORDS

eHealth; mHealth; behavior; neoplasm; mobile phones

Introduction

Background

The World Health Organization describes electronic health (eHealth) as a cost-effective and secure way to use information and communication technologies for health [1]. eHealth broadly encompasses the provision of health care services, education, allows surveillance, and the development of knowledge and research through technology [1]. Mobile health (mHealth) is a subset of eHealth, where mobile devices support the delivery of medical and public health care to individuals and populations [2]. In 2007, the number of individuals using the internet around the world was approximately 1.3 billion [3]. Furthermore, in 2018, this number had trebled to 3.9 billion people with internet access (51.2% of the world’s population) [3,4]. The use of mHealth in the delivery of health interventions is increasing worldwide because of the rapid growth of internet use and leaps in technological advancements. Its potential to target previously hard-to-reach populations and the need for innovative approaches to deliver health promotion and interventions in the face of aging populations and health care budget constraints make it an attractive delivery method [2,5-8].

Behavior change interventions can be defined as “coordinated sets of activities designed to change specified behaviour patterns” [9]. The effectiveness of digital behavior change interventions to promote behavior change is likely to be dependent on a complex interplay of factors, which are still in their infancy in terms of understanding. A systematic review of behavior change interventions delivered via the internet found that more extensive use of underlying behavior change theory and utilization of more behavior change techniques were important factors in overall intervention effectiveness [10]. Although eHealth interventions offer the promise of enhancing health care to populations in rural and regional settings and overcoming some of the challenges associated with accessing traditional health care delivery modes, there has been concern that in some segments of the community, provision of eHealth may exacerbate already prominent inequalities [11]. Of particular concern are individuals who have low health literacy, access to technology, and familiarity and confidence in the use of technology [11-14].

eHealth interventions, focusing on behavior change, are being increasingly used in patients with cancer and survivors. In 2018, cancer rates around the world increased to 18.1 million cases per year with 1 in 6 women and 1 in 5 men receiving a diagnosis in their lifetime [15]. Cancer survivorship also increased, with 43.8 million people surviving up to 5 years [15]. Many of these individuals live with long-term treatment side effects, including cardiac dysfunction, functional decline often precipitated by chronic pain and fatigue syndromes, obesity, diabetes, osteoporosis, premature menopause, neurocognitive deficits, and risks associated with primary recurrence and second cancers [16-18]. There are well-established guidelines and recommendations to modify risks associated with physical activity and healthy eating behavioral patterns post cancer diagnosis [19-21]. Studies of behavior change interventions in this population often use a guideline as a basis to improve physical activity and healthy eating behaviors. A recent review of self-guided technology-supported nutrition and physical activity interventions in adults with cancer found benefits in physical activity and fatigue with some modest effect on dietary behaviors and health-related quality of life [22]. Another review looked at telephone, print, and web-based interventions and found that improvements were reported in 76% of the studies included for physical activity, dietary behaviors, or weight [23]. However, these reviews did not contrast the relative effectiveness of different approaches to delivering eHealth behavior change interventions. There are 3 important categories that eHealth interventions can be described: synchronous, asynchronous, and combined. Synchronous eHealth interventions are delivered via real-time interactions between the person and health care provider, encompassing face-to-face contact through teleconferencing equipment, telephone (telehealth), and live chat via web-based [24-26]. Asynchronous techniques include store-and-forward methods such as email and automated messaging systems without a live interaction component [24-26]. Combination approaches use both synchronous and asynchronous elements. These distinctions are important because they impact how health services need to structure and staff services that use these approaches and dictate whether or not health services have to rely on technology-based platforms to enable delivery.

Objectives

There is a need to characterize effective eHealth interventions to enable improved design and implementation of evidence-based cancer health behavior change interventions, which will translate into the ability to scale up to affect health behavior change in a wide range of health promotion and health care management situations. This systematic literature review seeks to compare synchronous with asynchronous delivery modes and contrast their impact on behavior change and quality of life outcomes in adult patients with cancer or survivors. It also seeks to examine whether the degree of behavior change influences the amount of change in quality of life and to describe the behavior change theories and techniques used in the field.

Methods

In conducting this review, we recognized that increased access to mobile internet technologies, and increased availability of health information on the web, combined with changing behaviors in accessing health information [27,28] means that people who previously may have relied on health professionals to provide them with information are now better enabled and more likely to find information on their own. We chose to restrict study selection to being from a contemporary period (from 2007 onward) under the justification that the substantial increase in availability of information about health behaviors and managing the lifestyle consequences of cancer through the internet has evidently changed how people may seek and find information about their health issues and respond to behavior change interventions delivered by distance.

Reporting Guidelines

This systematic review follows the Cochrane Collaboration's Handbook of Systematic Reviews of Interventions [29]. This systematic review was registered with The International Prospective Register of Systematic Reviews (PROSPERO) CRD42018103855. The methods used in this systematic review are in line with the preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidelines. A completed copy of the PRISMA checklist is attached (Multimedia Appendix 1).

Search Strategy

A three-part search strategy was used to identify studies that met the following inclusion criteria: (1) we searched electronic bibliographic databases for published work including SCOPUS, Ovid MEDLINE, Excerpta Medica dataBASE (EMBASE), Cumulative Index to Nursing and Allied Health Literature (CINAHL) Plus, PsycINFO, and Cochrane CENTRAL; and (2) we searched the reference lists of the primary studies included in the review. We undertook a second phase of study identification where we completed a hand search of the *Journal of Medical Internet Research Cancer* and expanded the database search to include PubMed with the additional search term "website" (Multimedia Appendix 2).

Search Terms

Example Search Strategy

Mobile app*, electronic mail, internet, mhealth, mobile health, ehealth, electronic health, telehealth, telemedicine, telenursing, telemonitoring, telerehabilitation, telephone, cell* phone*, cell* telephone*, mobile telephone, mobile phone*, smartphone*, email*, electronic messag*, electronic mail, text messag*, short messag* service*, SMS, MMS, interactive voice response, multimedia, web-based, automat* reminder*, videoconferenc*, online*

AND

behavio* chang*, health behavio* chang*, behavio* theory, behavio* modifi*, health promotion

AND

NOT child*, adolescen*, teen*, preschool*, infant*, toddler*

AND

specific validated database filters for randomised controlled trials

AND

English language, 2007 to current

The asterisk truncates the search term so that alternative terms are also identified (eg, behavio* will find behavior and behavioral).

Titles and abstracts were retrieved using the search strategy. Reference lists were then exported into Clarivate Analytics EndNote X8 and duplicates were removed. References were then exported into Covidence where further *deduplication* occurred. Abstracts were reviewed by the author (KF), and an independent author (MS) performed a parallel review. The overall review was limited to the population of people with cancer when the abstract screening resulted in over 400 studies to be reviewed. Papers where reviewers disagreed on the rating of eligibility criteria were re-examined and discussed to reach consensus. The full-text papers were then retrieved and independently assessed by 2 reviewers (KF and MS). Hand searching for primary studies included in the review resulted in a further 23 studies for review.

Selection Criteria

Studies were limited to those relating to cancer patients or survivors, published with one or more search terms, were subjected to peer review, published in the English language, involved human adult subjects aged 18 years and over, and dated from 2007 to March 2019. The primary intervention was delivered through an eHealth delivery method such as telephone or internet, either asynchronous or synchronous or combined interventions against a control (including usual care or wait list control or no intervention), random assignment of participants to treatment or comparison groups, and a measure of health behavior change must have been taken after the intervention. Studies were excluded if there was any face-to-face component, as we sought to examine interventions purely delivered via eHealth approaches.

Outcomes

The Kirkpatrick model [30] is an internationally recognized tool to evaluate the effectiveness of training interventions. In this review, training is defined as the health behavior change intervention. We sought to examine relationships between the intervention delivery mechanism (synchronous or asynchronous) and the 4 different levels of the Kirkpatrick model. The 4 levels of the model are as follows:

- Level 1 (reaction): This is how participants responded to the training. This review focuses on participant user engagement metrics to assess reaction.
- Level 2 (learning): This is the content learned from the training provided and is usually knowledge tests completed pre- and postintervention.
- Level 3 (behavior): This level examines the participants' application of learning. We review the participants' ability to translate the intervention into health-related behavior change.
- Level 4 (results): This is the degree to which targeted outcomes have occurred as a result of the training provided. This review looks at health outcomes as described by previous studies.

This review examined how the different delivery mechanisms of behavior change intervention not only impact the individual levels of this model but also impact the interaction between these levels. We did not extract data related to the second level of the Kirkpatrick hierarchy (learning) as behavior change interventions do not necessarily require new information to be learned to change behaviors (this is a minor limitation in applying an educational framework in a behavior change context).

Primary Outcomes

The primary outcome for this review was defined as a change in health behaviors (eg, increases in moderate-to-vigorous physical activity or fruit and vegetable consumption) from baseline to the conclusion of the active intervention. This was chosen as it was most likely to be the time of greatest compliance with the intervention.

Secondary Outcomes

The secondary outcomes for this review were engagement in the eHealth interventions and changes in quality of life, fatigue, anxiety, and depression. We sought to measure the proportion of participant initiation of the intervention, the frequency of intervention delivery per week, and duration of intervention delivery in minutes. Changes in quality of life, fatigue, anxiety, and depression were measured relative to the measure taken as close to the conclusion of the active intervention as possible, not after any period of follow-up.

Other Descriptive Data

We sought to describe the behavior change theories used and behavior change techniques employed in eHealth interventions.

Data Extraction

One (KF) and either of the 2 reviewers (MS or KH) independently extracted data including study identifiers, study design, population characteristics, consent and retention rate, intervention, behavior change and health outcomes of interest, behavior change theory and techniques used, intervention participation rates, control condition details, length of follow-up, and information to complete a risk of bias assessment into a standardized data extraction table. Data were extracted on the key outcomes (objective or self-reported) defined by the study, including those that reported on multiple behavior change outcomes (eg, diet and physical activity). Where this was not clear, 2 reviewers independently chose the most relevant behavior change outcome to the study. Where data were incompatible for meta-analysis, authors were contacted twice via email, and 2 out of 9 authors who were contacted responded with forthcoming information. Engagement data were also extracted by 2 reviewers.

Risk of Bias Assessment

Two review authors (KF and KH) independently assessed the risk of bias for randomized controlled trials (RCTs) using the Cochrane Collaboration's tool for assessing bias [31]. Differences of opinion were discussed and agreed upon between the 2 reviewers.

Analysis

Behavior Change and Health Outcomes

Data were separated into groups of similar outcomes to perform pooled random effects meta-analysis using standardized mean difference (SMD) of summative level data. Standardized effect sizes were considered small at 0.2, moderate at 0.5, and large at 0.8 [32]. A random effects metaregression analysis was undertaken to determine if the mode of delivery influenced the amount of behavior change observed. The SD of the effect size was imputed for one study [33], as variance data were not available. The imputation method was used to borrow a particular value from other studies using the same outcome measure [34]. The functional assessment of chronic illness therapy-fatigue, used in 3 studies [35-37], was transformed by multiplying the mean values by −1 to ensure that all scales were in the same direction [29]. For visual representation and ease of interpretation, fatigue, depression, and anxiety were transformed to allow lower scores to represent improved outcomes.

Initiation of Engagement Data

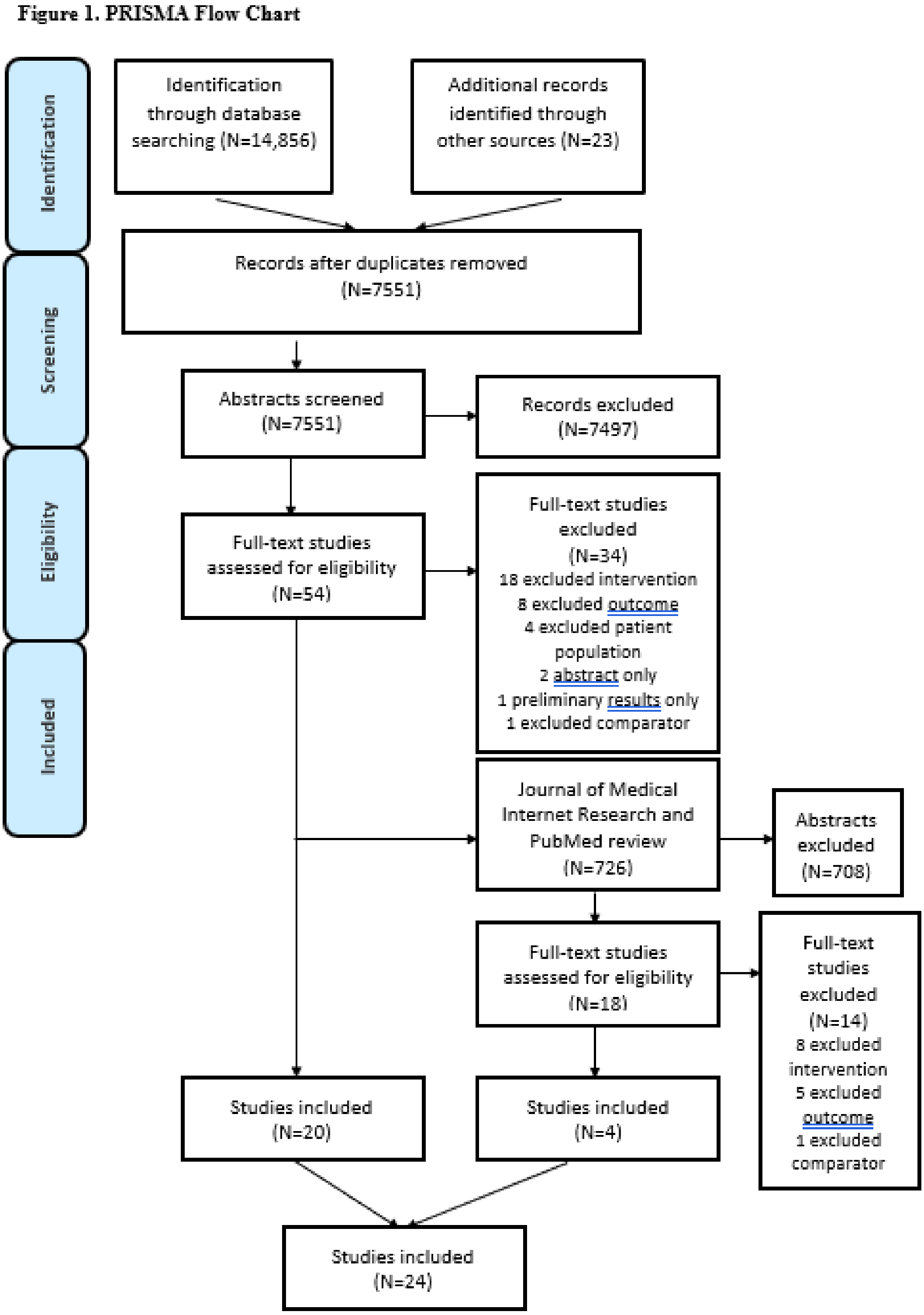
Variances of raw proportion data were transformed using the Freeman-Tukey arcsine square root transformation [38]. The DerSimonian-Laird random effects model [39] was then used to pool these transformed proportions to provide a measure of the extent of between-study heterogeneity. The Wilson score was then used to provide CIs for the pooled estimate [40]. A random effects plot was then created. Random effects metaregression analyses were then performed to determine whether any of the different delivery modes influenced the amount of engagement initiation.

Results

Summary

The literature search identified a total of 15,582 studies with a further 23 studies identified through reference list searching of relevant papers. After EndNote X8 and Covidence deduplication, a total of 8259 studies were screened, with 72 s remaining for full-text screening. A total of 24 studies were included in this review; 11 studies were delivered via synchronous and asynchronous methods, respectively, with 2 studies using a combined approach (Figure 1). The studies were predominantly delivered in the United States with 13 studies [37,41-52]; 3 in Australia [35,36,53], South Korea [33,54,55], and the Netherlands [56-58]; and 1 in Canada [59] and France [60]. The total sample size was 4583, ranging from n=18 to n=641. There were a range of different cancer types targeted during the study interventions: 7 unspecified cancer types [45,49,50,52,55-57,60]; 6 breast cancer [33,35,42,51,54,61]; 2 colon or colorectal [36,43]; 2 breast, prostate, or colorectal [44,59]; 2 melanoma [41,53]; and 1 each of breast/colon or rectal [37], colorectal or prostate [58], prostate [46], and urothelial cell carcinoma [47]. The majority of studies focused on posttreatment (also reported as survivors) with 17 studies [33,36,37,41,43-45,49-56,59,61], 6 studies included patients during active cancer treatment (2 used survivors and active cancer treatment) [35,42,47,57,58,60], and 1 study included patients undergoing surveillance for their cancer [46]. Of the 24 studies included in this review, 19 provided data on a measure of physical activity behavior change [33,35-37,42-45,48-50,54-57,59] and 9 studies provided data on a measure of dietary behavior change [33,36,42,44-47,54,56]; 2 studies provided data on skin self-examination during melanoma-related interventions [41,53]; 2 studies provided data on smoking cessation [52,56], whereas another study provided information on alcohol intake [36]. Interventions delivered via synchronous modes included telephone, Skype, and videoconferencing, whereas those delivered via asynchronous modes used combinations of custom or existing websites and mobile apps, with short messaging and email. Studies that used combined methods used web-based intervention and an online moderated forum, and telephone and SMS text messaging. Many studies in any of the delivery method intervention groups used adjunct features, including pedometers and written workbooks (Multimedia Appendix 3).

Figure 1. Preferred reporting items for systematic reviews and meta-analysis flowchart.



Behavior Change Techniques

The use of behavior change techniques varied significantly between studies; studies used between 2 and 17 (median 8) different techniques. The most popular behavior change techniques used across the studies were goal setting (behavior; 87%) and self-monitoring of behavior (85%), information about health consequences (70%), problem solving (67%), action planning (62%), feedback on behavior (61%), and instructions on how to perform behavior (61%). Out of the 93 hierarchically set behavior change techniques from the Michie et al [62] taxonomy, only 33 were used across the 24 studies included in this review (Multimedia Appendix 4).

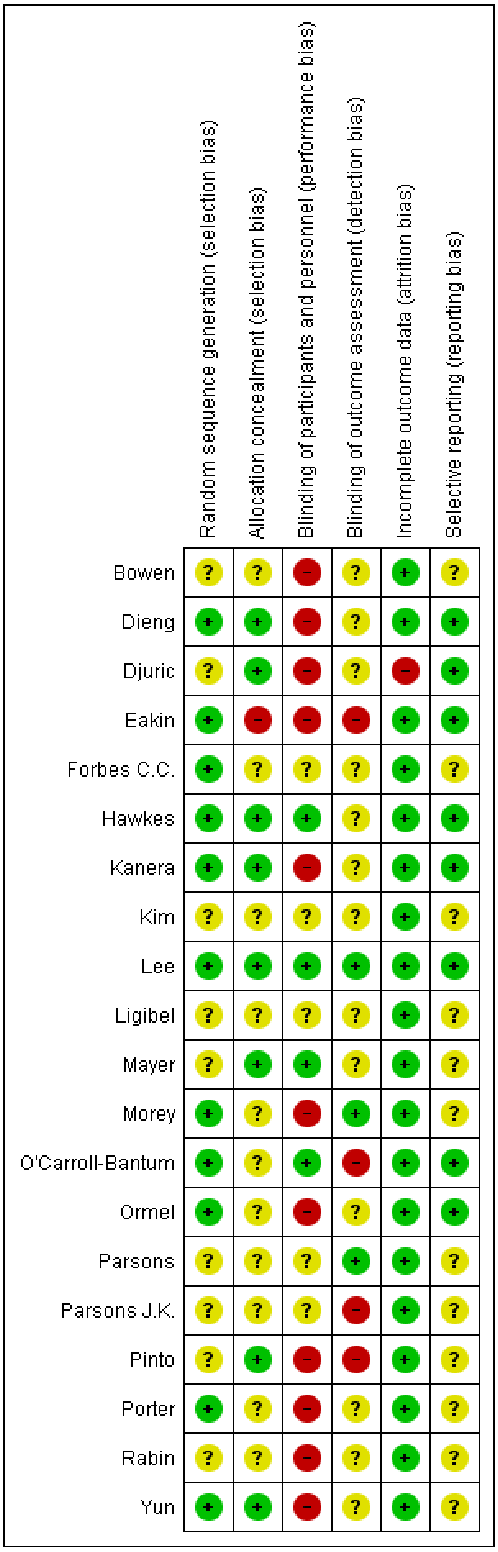
Behavior Change Theory

A total of 20 studies reported on the use of behavior change theory. Within the studies that reported on behavior change theory, social cognitive theory and the transtheoretical model of change were the 2 most popular theories used, with 60% and 39% of papers reporting on these, respectively. A total of 8 studies (33%) used more than one theory to underpin their interventions (Multimedia Appendix 4).

Risk of Bias Assessment

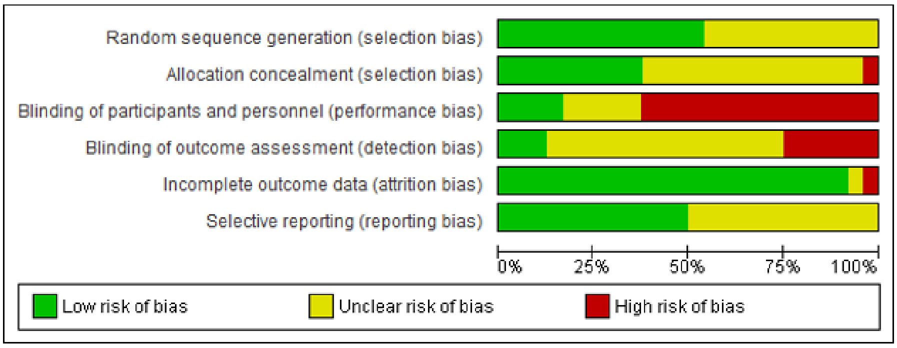
The outcome of the risk of bias assessment is presented in Figures 2 and 3. Attrition bias was considered to be low across most studies. There was a high degree of risk associated with blinding of personnel, participants, and outcome assessment. The reporting of study-related processes was highly variable, which led to many areas of the risk of bias being assigned as unclear.

Figure 2. Risk of bias summary of the included randomized control trials using the Cochrane Collaboration risk of bias tool.



	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)
Bowen	?	?	−	?	+	?
Dieng	+	+	−	?	+	+
Djuric	?	+	−	?	−	+
Eakin	+	−	−	−	+	+
Forbes C.C.	+	?	?	?	+	?
Hawkes	+	+	+	?	+	+
Kanera	+	+	−	?	+	+
Kim	?	?	?	?	+	?
Lee	+	+	+	+	+	+
Ligibel	?	?	?	?	+	?
Mayer	?	+	+	?	+	?
Morey	+	?	−	+	+	?
O'Carroll-Bantum	+	?	+	−	+	+
Ormel	+	?	−	?	+	+
Parsons	?	?	?	+	+	?
Parsons J.K.	?	?	?	−	+	?
Pinto	?	+	−	−	+	?
Porter	+	?	−	?	+	?
Rabin	?	?	−	?	+	?
Yun	+	+	−	?	+	?

Figure 3. Risk of bias graph.



Behavior Change: Physical Activity

Of the 19 studies in this review that presented data on physical activity as a behavior change outcome, 15 studies were included in this meta-analysis [35-37,42,44,45,48-51,54,57-60] and a further 4 could not be included as the data were presented in a format incompatible with meta-analysis [33,43,55,56]. There was a finding in favor of eHealth interventions (SMD 0.34; 95% CI 0.21 to 0.48) for increasing physical activity behaviors (Figure 4).

When analyzed by delivery method, the intervention delivery mode of synchronous, asynchronous, or combined did not impact the overall positive outcome effect (Table 1).

Figure 4. Effect of electronic health interventions compared with a control on physical activity interventions, analyzed using standardized mean difference. SMD: standardized mean difference.

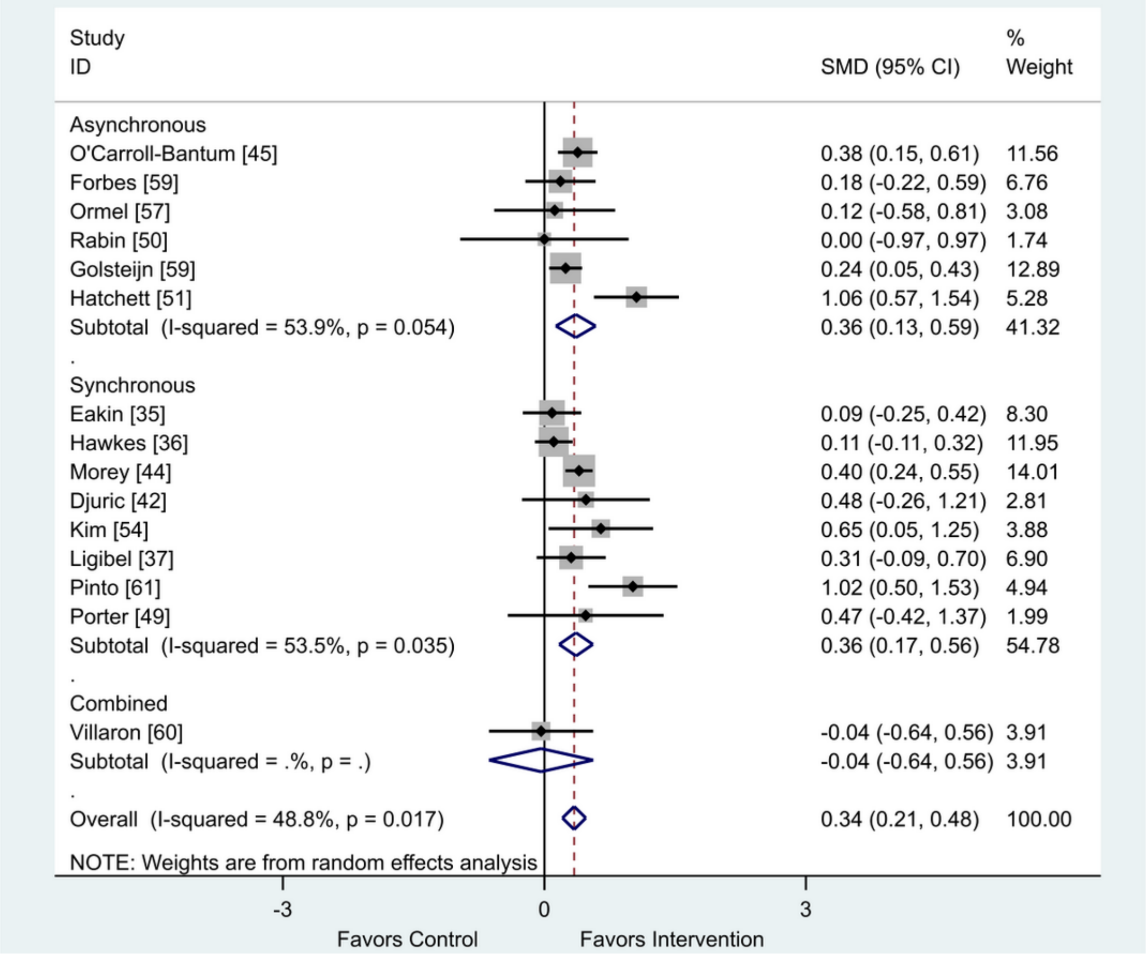


Table 1. Metaregression comparing the effect of synchronous versus asynchronous versus combined electronic health interventions on physical activity behavior change, analyzed using standardized effect size.

Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	−0.01	−0.38 to 0.35	.95
Combined	−0.41	−1.27 to 0.45	.32
Constant	0.37	0.13-0.61	.005

The 4 studies examining asynchronous interventions provided data that were presented in a format incompatible with meta-analysis. One reported a significant increase in moderate physical activity in their intervention group (*P*=.04); however, the authors reported that this did not remain significant after controlling for multiple testing [56]. Another study reported that the intervention increased the proportion of participants who undertook moderate intensity aerobic exercise for at least 150 min per week to a greater extent than the control (*P*=.001) [33]. These outcomes contrasted with 2 other studies that reported no effect in improving the secondary outcome of metabolic equivalent of task (*P*=.48) [55] and the proportion of participants who undertook moderate-to-vigorous physical activity (*P*=.12) [43].

Behavior Change: Dietary

A total of 9 studies included in this review provided data on diet as a behavior change outcome; 6 studies were included in this meta-analysis [36,42,44,45,47,54] and a further 3 could not be included as the data were presented in a format incompatible with meta-analysis [33,46,56]. There was a finding that diet interventions delivered via eHealth can improve behavior change compared with control (SMD 0.44; 95% CI 0.18 to 0.70; Figure 5).

When analyzed by delivery method, the intervention delivery mode of synchronous or asynchronous did not impact the overall positive outcome effect (Table 2).

Figure 5. Effect of electronic health interventions compared with control on dietary interventions, analyzed using standardized mean difference. SMD: standardized mean difference.

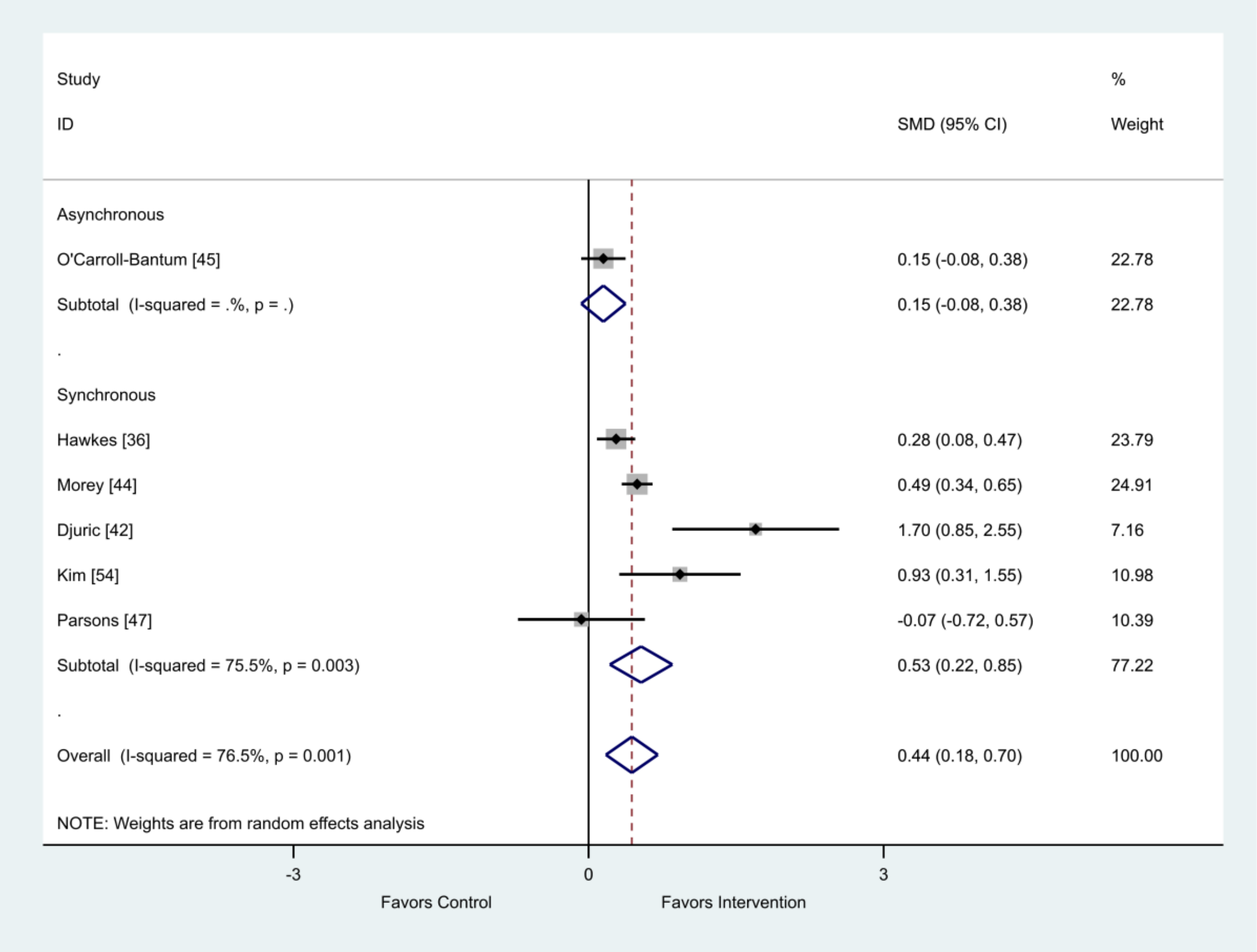


Table 2. Metaregression of synchronous and asynchronous dietary interventions on behavior change outcome, analyzed using standardized effect size.

Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	−0.44	−2.18 to 1.29	.52
Constant	0.60	−0.17 to 1.36	.10

In total, 3 studies provided data incompatible with the meta-analysis. The 2 studies using asynchronous delivery approaches reported significantly higher vegetable consumption in participants accessing a web-based guide (*P*=.02), which the authors reported did not remain significant after accounting for multiple testing [56], whereas dietary quality index was greater in the intervention compared with the control in the other study (*P*=.001) [33]. One synchronous intervention reported a significant increase in alpha-carotenoid concentrations compared with controls (*P*<.05) [47].

Other Behavior Change Outcomes

A total of 5 studies provided data on other primary behavior change outcomes, including smoking cessation, alcohol intake, and skin self-assessment. A study examining asynchronous interventions reported no effect on smoking cessation (*P*=.28; odds ratio [OR] 2.61) [56]. Similarly, no improvement in smoking cessation rates was found in a study using a combined delivery mode approach [52]. A separate trial using an asynchronous delivery mode to encourage skin self-examination was effective (OR derived from percentage data provided 1.90, 95% CI 2.23-2.94) [41]. Conversely, another study examining the effect of an intervention (delivered synchronously) to encourage skin self-examination reported a decrease in the reported rate of skin self-examination in the intervention group compared with control (adjusted between-group difference −0.13; 95% CI −0.4 to 0.2; *P*=.40); however, this outcome was measured at 6 months follow-up rather than immediately post intervention at 1 month. This study also presented information about melanoma-related knowledge change and was the only study included in this review that presented any data on knowledge change. They found an improvement in melanoma-related knowledge in the intervention group at 6 months using an adjusted between-group difference (1.7; 95% CI 0.8-2.6; *P*<.001) [53]. The final study examining a synchronously delivered intervention revealed no significant difference between groups on alcohol intake (grams per day; *P*=.26) [36].

Health Outcome: Quality of Life

There were 13 studies that provided data on quality of life [33,35-37,42-44,49,54,55,58-60]; 2 meta-analyses were performed as data were presented as a combination of final scores and change scores. Of the 7 studies that provided final scores, there was a favorable impact of the intervention for the synchronous (SMD 0.25; 95% CI −0.36 to 0.87) and combined (SMD 0.35; 95% CI −0.25 to 0.95) eHealth interventions on quality of life. The 3 studies using an asynchronous mode for intervention delivery reported no improvement in quality of life compared with control conditions (SMD 0.01; 95% CI −0.15 to 0.17; Figure 6).

Of the 5 studies that provided change scores, there was a favorable impact of the intervention for both asynchronous (SMD 0.12; 95% CI 0.01-0.23) and synchronous (SMD 0.26; 95% CI 0.02-0.49) eHealth interventions on quality of life (Figure 7).

Figure 6. Effect of electronic health interventions compared with a control on quality of life interventions, analyzed using standardized mean difference (postintervention mean and SD). SMD: standardized mean difference.

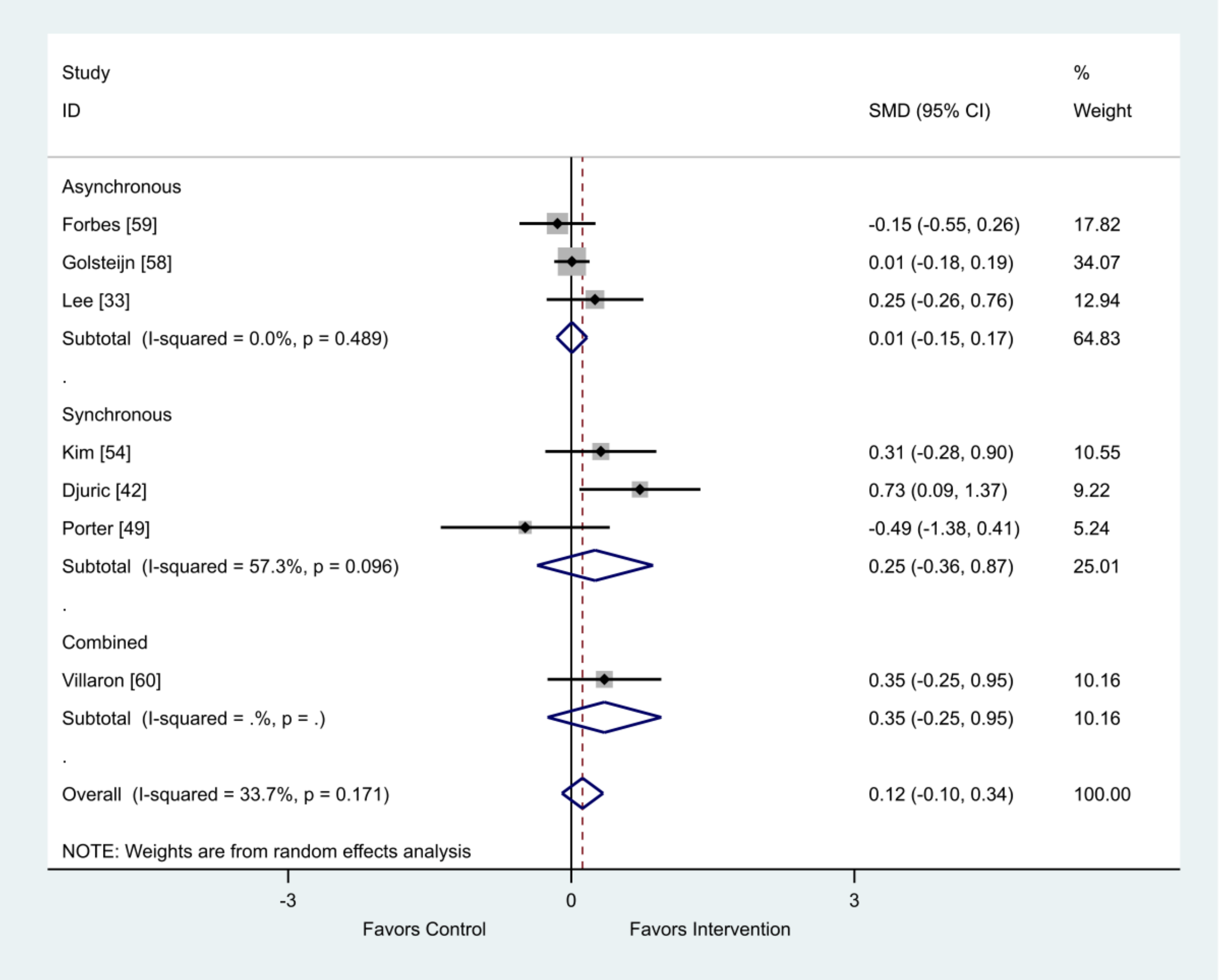
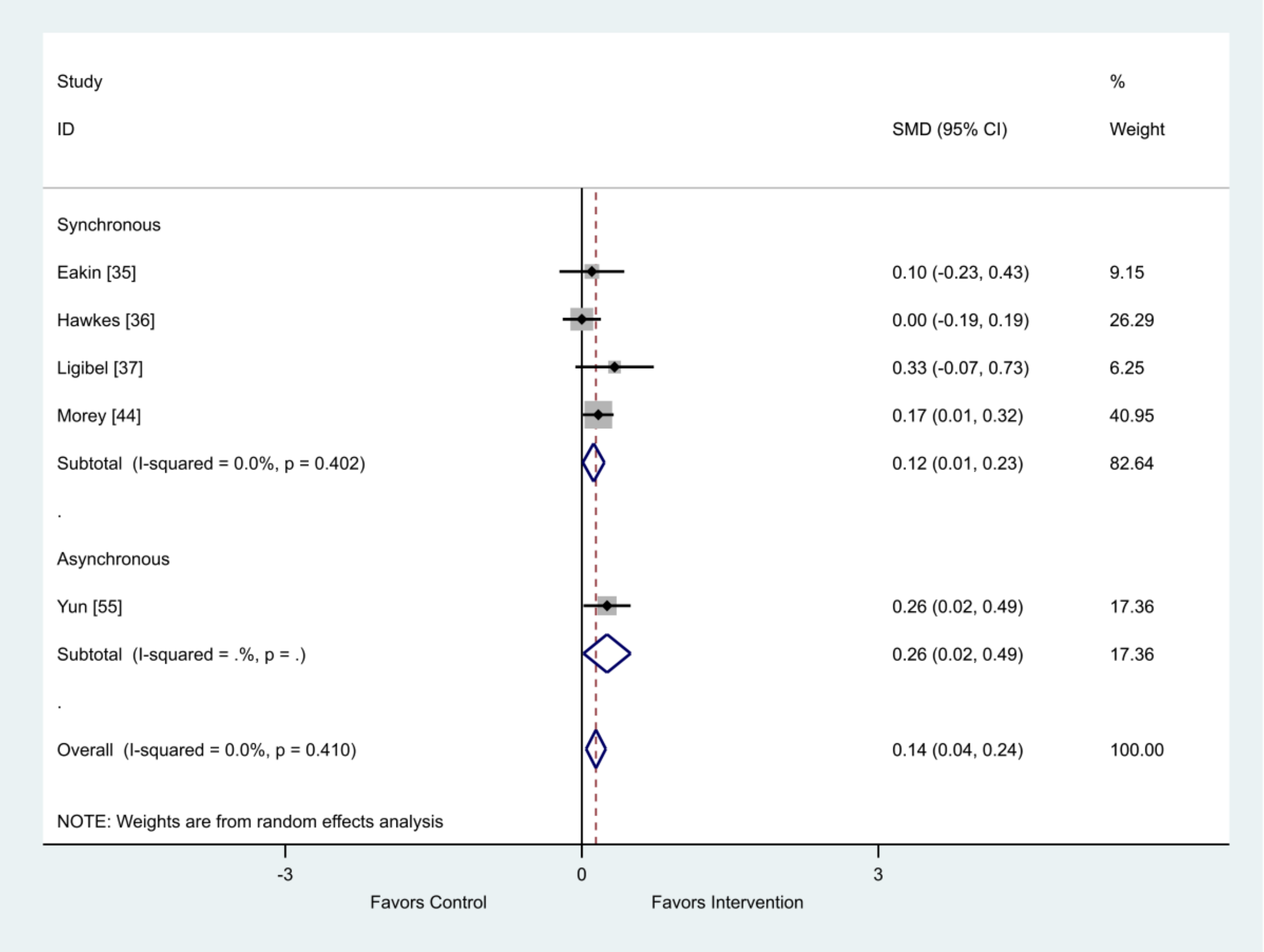


Figure 7. Effect of electronic health interventions with a control on quality of life interventions, analyzed using standardized mean difference (postintervention mean and SD change scores). SMD: standardized mean difference.



One study provided data that were incompatible with the meta-analysis. It was found that quality of life was not significantly different between the asynchronous intervention group and the no intervention control group over the intervention period. The intervention group baseline outcome score was 105 versus the postintervention outcome score of 109.1 and the control group baseline outcome score of 103.3 versus the postintervention outcome score of 106.5 [43].

When analyzed by delivery method, the intervention delivery mode of synchronous or asynchronous did not impact the overall positive outcome effect (Table 3).

When analyzed by delivery method, the intervention delivery mode of synchronous or asynchronous did not impact the overall positive outcome effect (Table 4).

Table 3. Metaregression of synchronous and asynchronous interventions on quality of life outcome, analyzed using standardized effect size.

Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	−0.31	−1.05 to 0.43	.31
Combined	0.04	−1.22 to 1.29	.94
Constant	0.31	−0.37 to 1.0	.27

Table 4. Metaregression of synchronous and asynchronous interventions on quality of life outcome, analyzed using standardized effect size (postintervention mean and SD change scores).

Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	0.14	−0.31 to 0.58	.40
Constant	0.12	−0.70 to 0.31	.14

Health Outcome: Fatigue

There were a total of 10 studies that presented data on fatigue [33,35-37,45,50,54,55,58,60,63]. Of the 6 studies using final scores, synchronous (SMD 1.03; 95% CI 0.41-1.66) and combined (SMD 0.23; 95% CI −0.37 to 0.83) interventions showed a favorable impact on fatigue. The asynchronous group reported no intervention effect on fatigue (SMD 0.03; 95% CI −0.18 to 0.24; Figure 8).

Of the 4 studies that provided change scores, both synchronous (SMD 0.19; 95% CI 0.03-0.34) and asynchronous (SMD 0.29; 95% CI 0.05-0.53) eHealth interventions showed a favorable impact on fatigue (Figure 9).

When analyzed by delivery method, the intervention delivery mode of synchronous, asynchronous, or combined did not impact the overall positive outcome effect (Table 5).

When analyzed by delivery method, the intervention delivery mode of synchronous or asynchronous did not impact the overall positive outcome effect (Table 6).

Figure 8. Effect of electronic health interventions with a control on fatigue interventions, analyzed using standardized mean difference (postintervention mean and SD). Lower scores indicate improved fatigue. SMD: standardized mean difference.

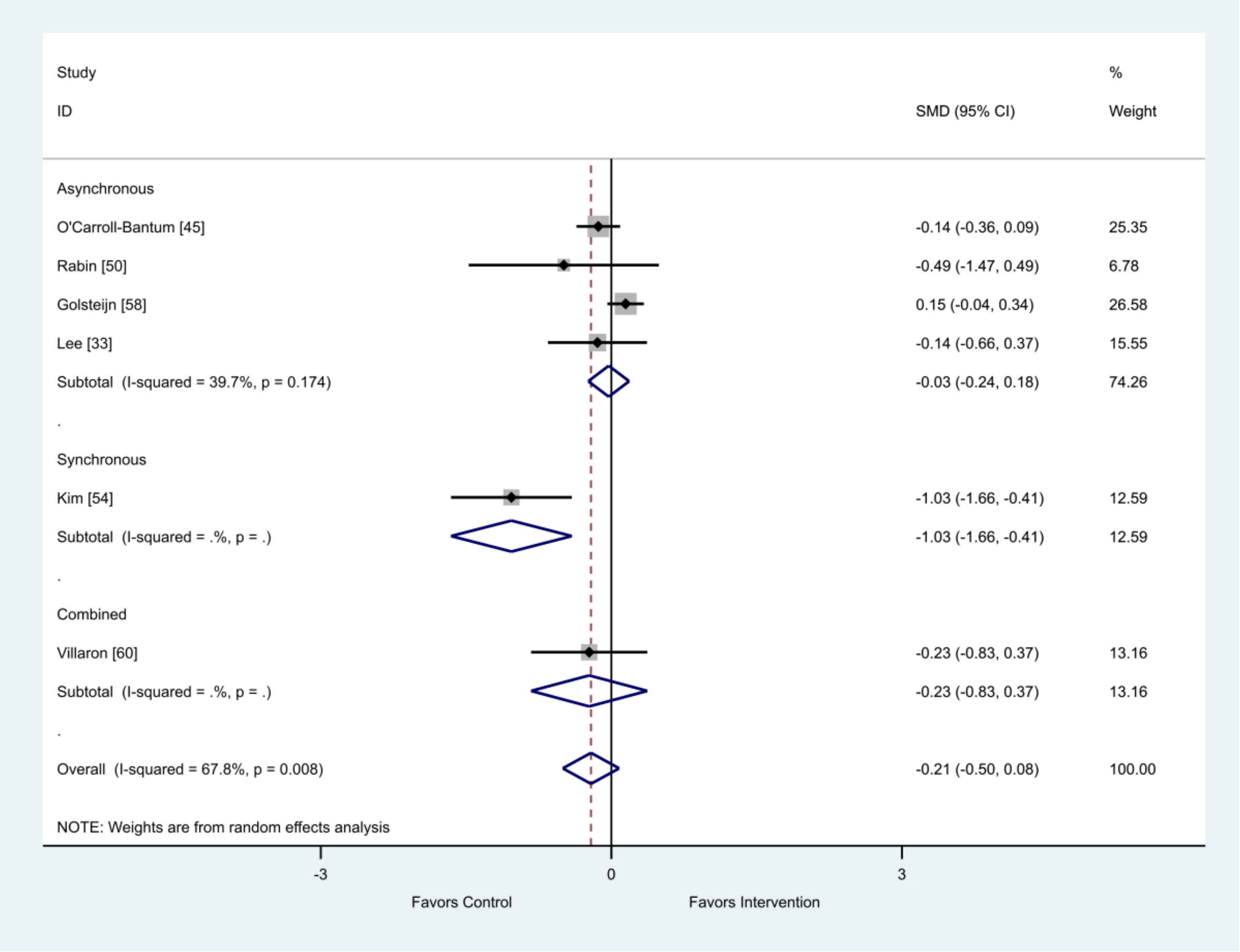


Figure 9. Effect of electronic health interventions with a control on fatigue interventions, analyzed using standardized mean difference (postintervention mean and SD change scores). Lower scores indicate improved fatigue. SMD: standardized mean difference.

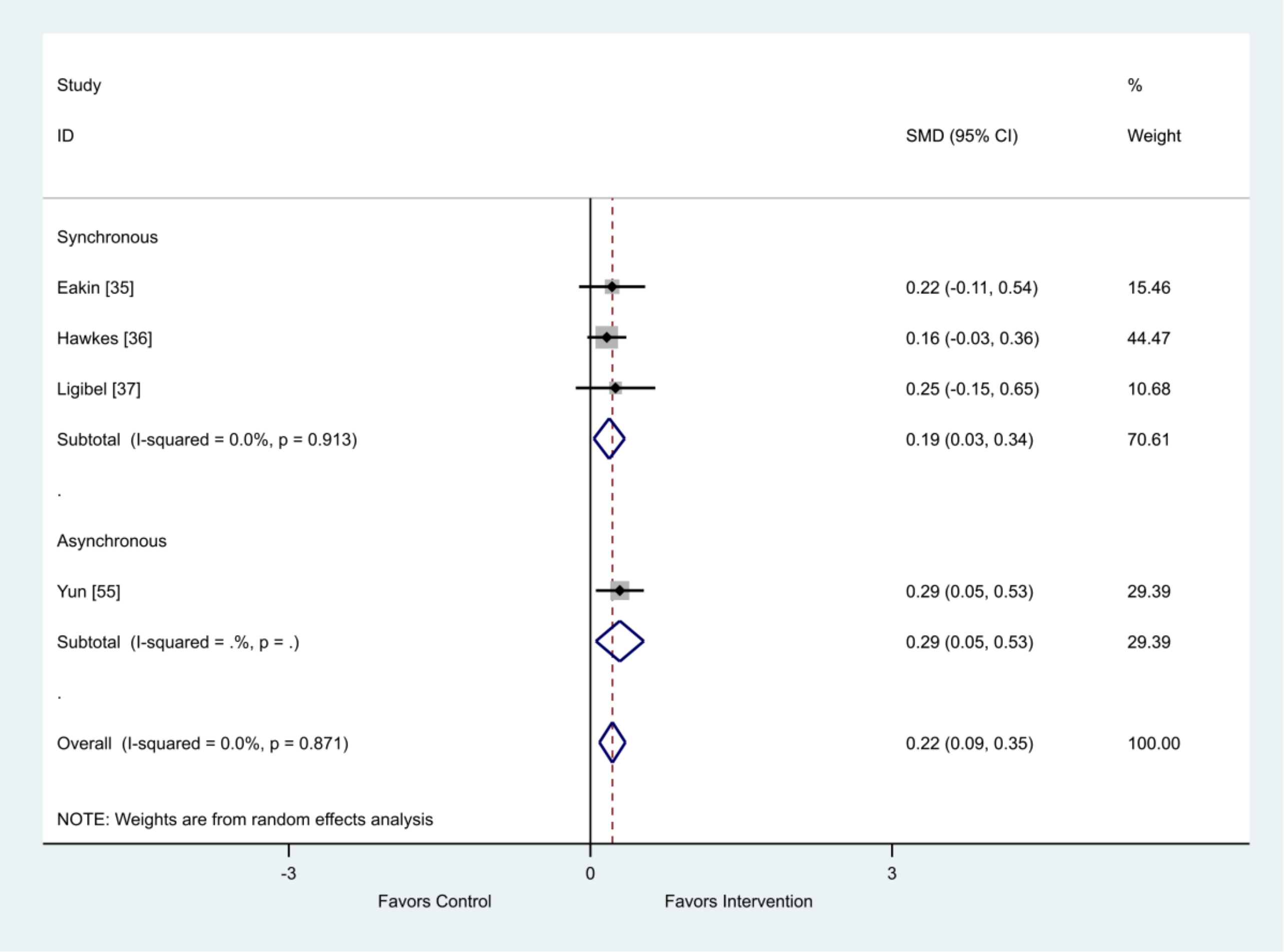


Table 5. Metaregression of synchronous and asynchronous interventions on fatigue outcome, analyzed using standardized effect size.

Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	1.00	−0.16 to 2.17	.07
Combined	0.80	−0.74 to 2.35	.20
Constant	−1.03	−2.14 to 0.77	.08

Table 6. Metaregression of synchronous and asynchronous interventions on fatigue outcome, analyzed using standardized effect size.

Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	0.11	−0.51 to 0.73	.54
Constant	0.19	−0.15 to 0.53	.14

Health Outcome: Depression

There were a total of 6 studies that presented data on health outcome depression [33,45,53-55,58]. Of the 5 studies’ final scores, the synchronous intervention showed a favorable impact on depression (SMD 0.80; 95% CI 0.19-1.41). The asynchronous group reported no intervention effect on depression (SMD 0.11; 95% CI −0.01 to 0.22; Figure 10). The one that compared asynchronous study with a no intervention control that reported change scores found no intervention effect on depression (group difference −1.0; 95% CI −2.8 to 0.8; *P*=.40).

When analyzed by delivery method, the intervention delivery mode of synchronous or asynchronous did not impact the overall positive outcome effect (Table 7).

Figure 10. Effect of electronic health interventions with a control on depression interventions, analyzed using standardized mean difference (postintervention mean and SD). Lower scores indicate improved depression. SMD: standardized mean difference.

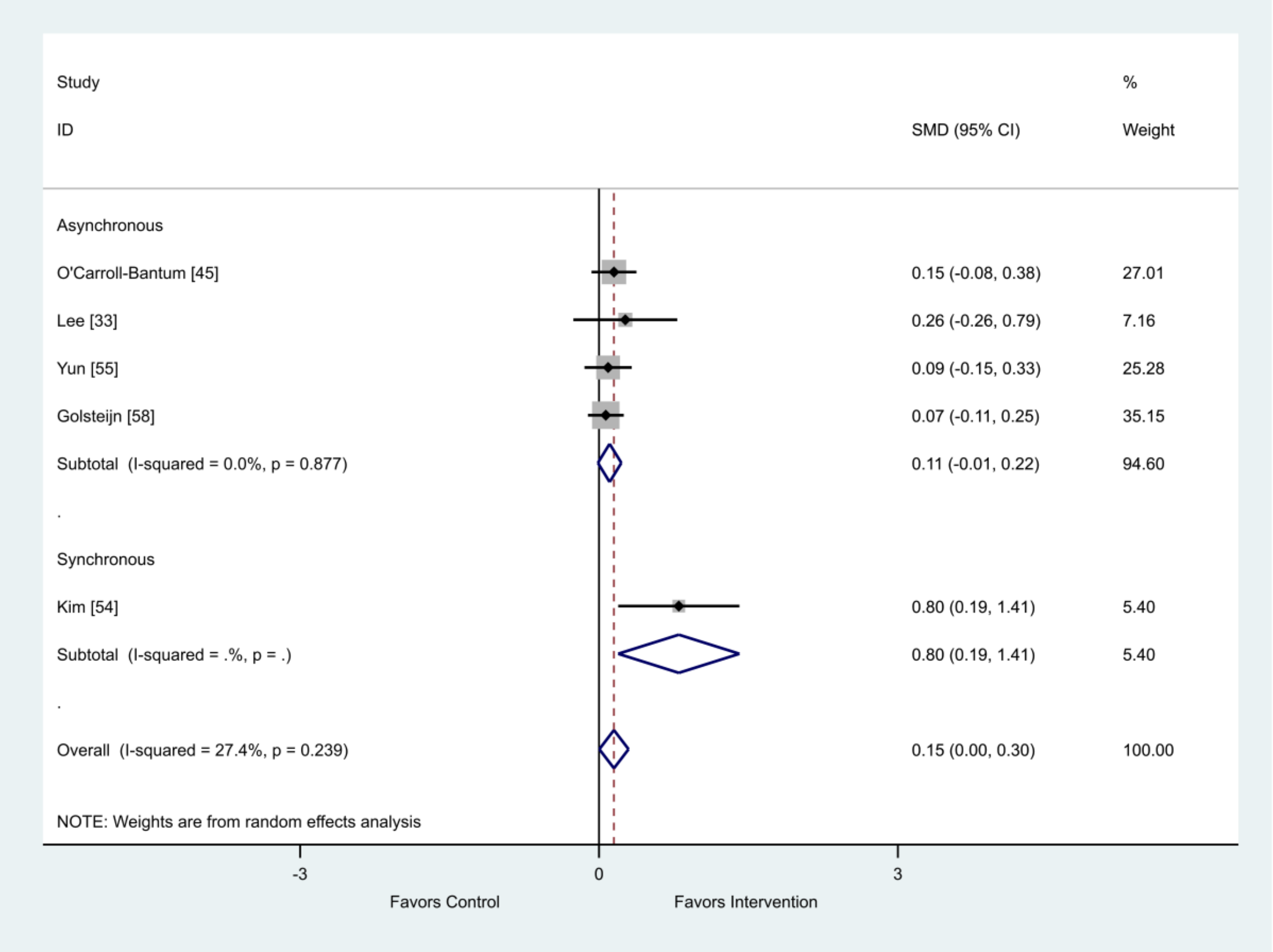


Table 7. Metaregression of synchronous and asynchronous interventions on depression outcome, analyzed using standardized effect size.

Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	−0.69	−1.70 to 0.32	.12
Constant	0.80	−0.19 to 1.79	.08

Health Outcome: Anxiety

There were 6 studies that provided health outcome data on anxiety [33,35,53-55,58].

Of the 4 studies that provided final scores, both synchronous (SMD 2.78; 95% CI 1.95-3.61) and asynchronous (SMD 0.74; 95% CI −0.01 to 1.48) modes showed a favorable impact on anxiety (Figure 11).

Of the 2 studies that provided change scores comparing synchronous interventions with usual care control groups, there was a very small favorable intervention effect on anxiety (SMD 0.15; 95% CI −0.09 to 0.40; Figure 12).

When analyzed by delivery method, the intervention delivery mode of synchronous or asynchronous did not impact the overall positive outcome effect (Table 8).

Figure 11. Effect of electronic health interventions versus control group on anxiety interventions, analyzed using standardized mean difference (postintervention mean and SD). Lower scores indicate improved anxiety. SMD: standardized mean difference.

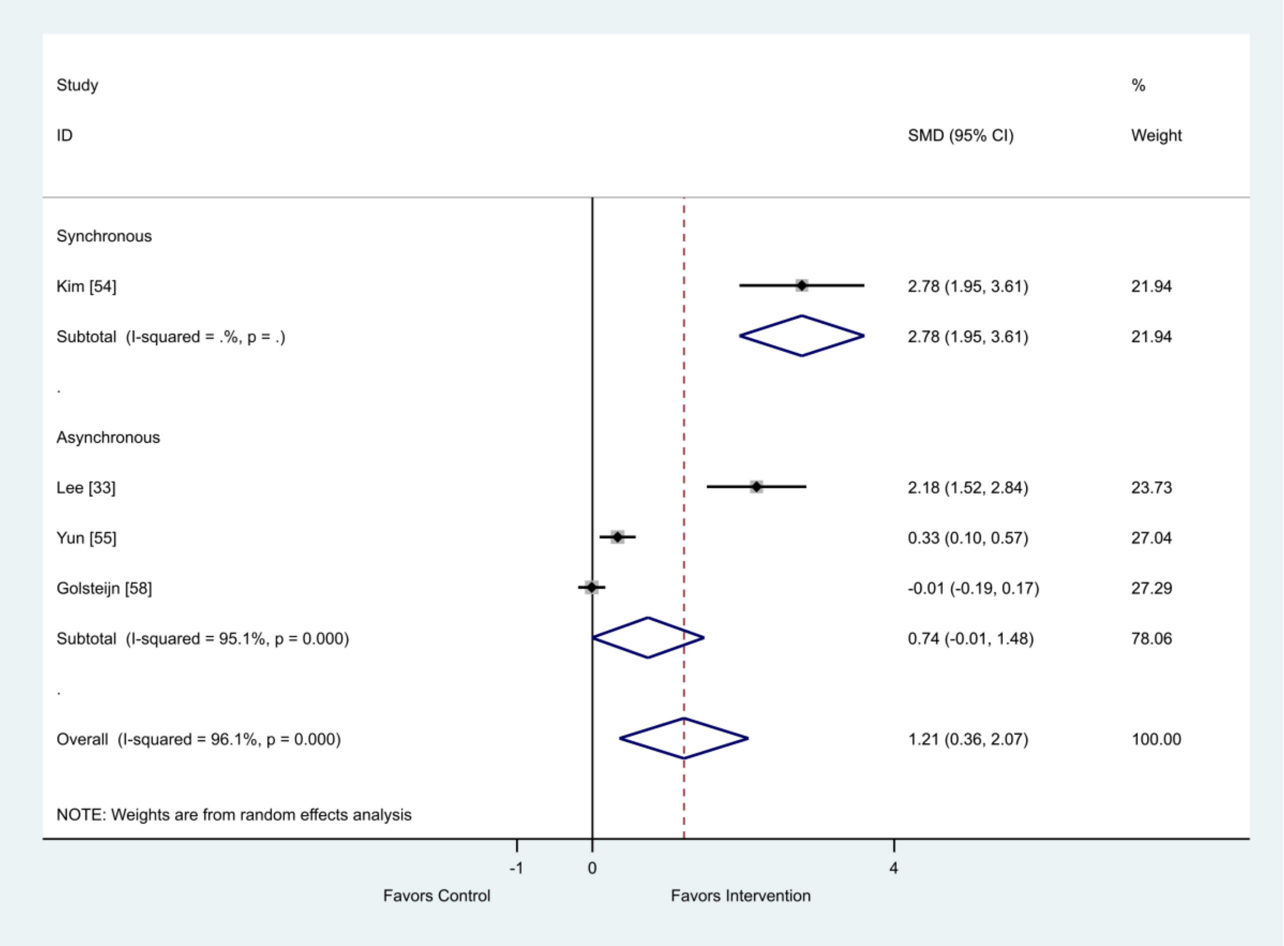


Figure 12. Effect of electronic health interventions with a control on anxiety interventions, analyzed using standardized mean difference (postintervention mean and SD change scores). Lower scores indicate improved anxiety. SMD: standardized mean difference.

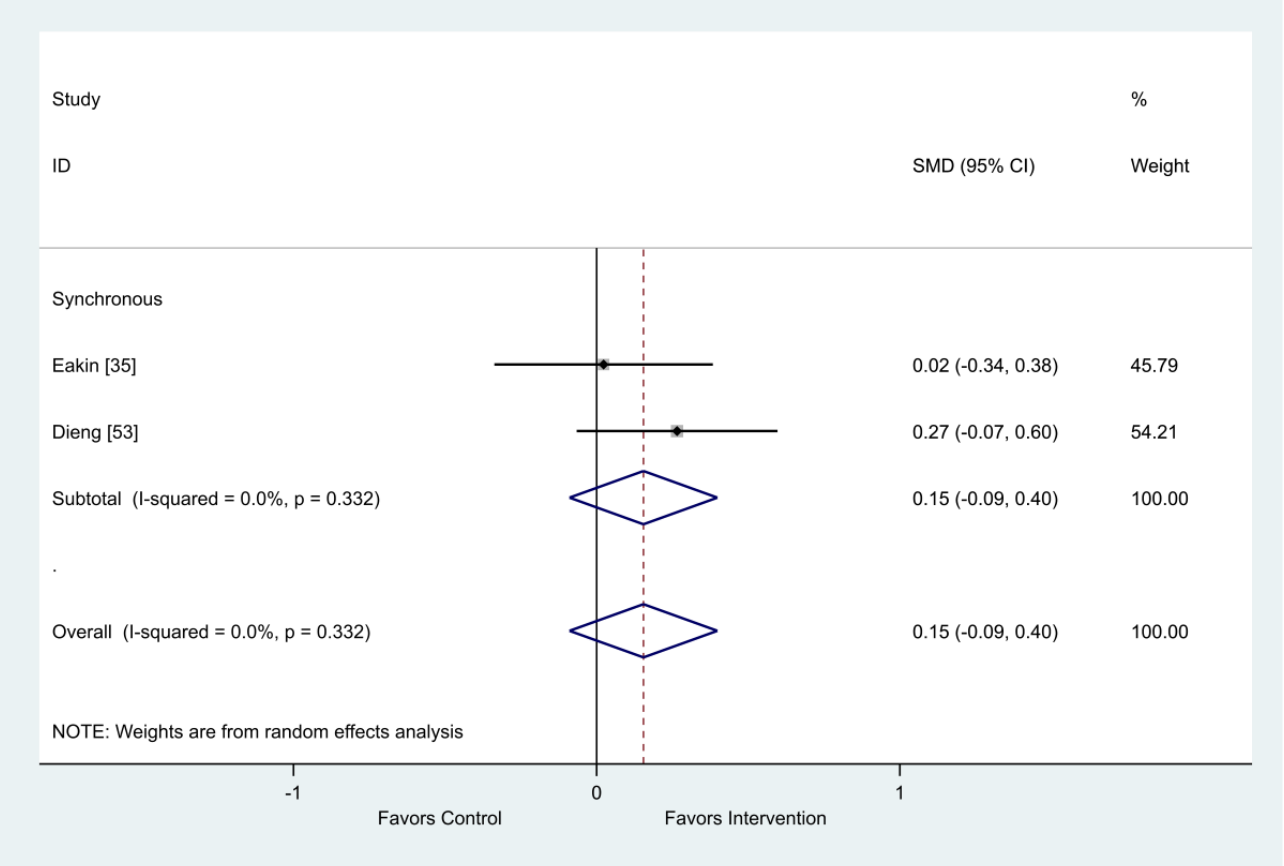


Table 8. Metaregression of synchronous and asynchronous interventions on anxiety outcome, analyzed using standardized effect size.

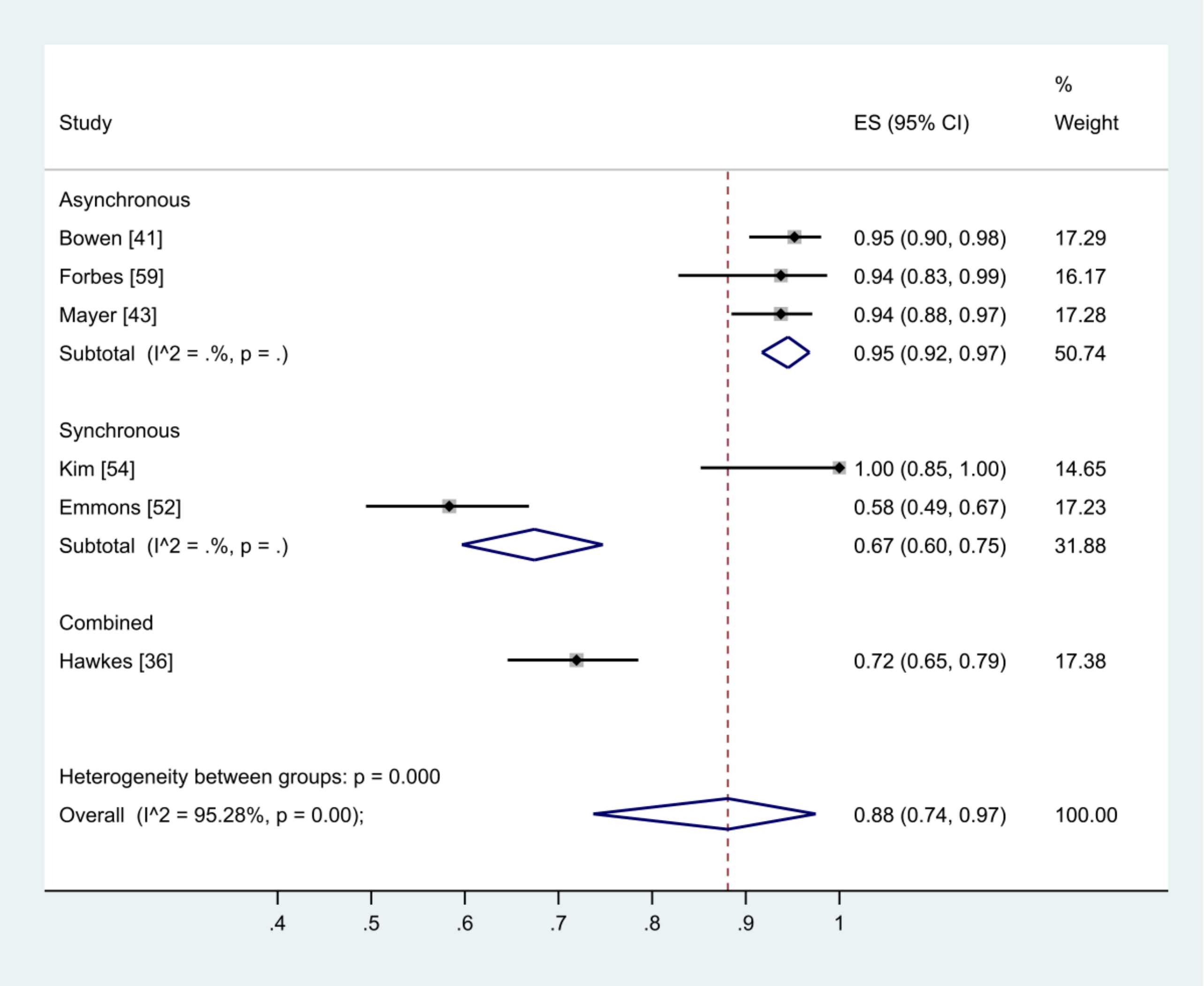
Standardized effect size	Coefficient	95% CI	*P* value
Asynchronous	−1.98	−7.96 to 4.01	.29
Constant	2.77	−2.47 to 8.02	.15

We were unable to perform metaregression to examine the association between mode of delivery and change in anxiety, as there was only 1 delivery mode (synchronous) included in this meta-analysis.

Initiation of Engagement Data

A total of 5 studies did not provide any information about engagement or participation in their intervention [46,48,51,57,58]. Only 6 studies provided data on the initiation of participants to their interventions [36,41,43,52,54,59]. Other data pertaining to participant engagement or participation in interventions are shown in Table 1. The pooled effect of the proportion of initiation engagement found that 88% of participants commenced the intervention (Figure 13). Of the 11 synchronously delivered interventions, 9 studies reported on completion of telephone or videoconference sessions. In all, 7 studies reported that participants completed between 62% and 100% of all planned intervention sessions [33,36,42,45,49,53]. One study reported that 79% completed >75% of telephone calls [35] and another reported a median 9 out of 10 to 11 planned telephone calls completed [37]. Comparison of engagement and participation was made particularly challenging in the asynchronous group due to the heterogeneous nature of the reporting. Studies presented a range of data that included initiation, content accessed, log-in averages, and intervention fidelity or adherence.

Figure 13. The effect of mode of delivery (synchronous or asynchronous) interventions on initiation engagement, analyzed using proportion.



Discussion

Principal Findings

The use of eHealth interventions improved physical activity, diet behaviors, quality of life, fatigue, depression, and anxiety in studies conducted across people who had been treated for a range of cancer types compared with control conditions. The overall impact appeared to be greater for the behavior change outcomes than the health outcomes. It did not appear that the mode of delivery (synchronous vs asynchronous vs combined) affected how much improvement the eHealth interventions generated. There was insufficient evidence to identify that the amount of behavior change was associated with the amount of change in health outcomes, although this may yet be identified as further studies report on both outcomes. These findings indicate that eHealth interventions may be beneficial for improving health behaviors and health outcomes when provided to cancer patients and/or survivors. However, there was one behavior change area where a negative impact on a health behavior (skin self-examination for melanoma) was reported, indicating that not all health behaviors may be improved from exposure to eHealth interventions.

At this stage, there is insufficient evidence to determine what delivery methods work for who, in what context, at what time. Therefore, the decision as to whether health services should provide synchronous or asynchronous eHealth behavior change interventions cannot yet be answered. None of the studies in this review compared synchronous with asynchronous eHealth approaches in the same population, using the same behavior change theory and techniques. A study that examined a telephone-delivered versus an internet-delivered weight loss intervention in cancer survivors found that the telephone group not only had greater losses in waist circumference (−0.75 vs −0.09; *P*=.03) but also had higher levels of engagement (80% completed calls vs 27% web log-ins) [64]. Other considerations, such as the cost-effectiveness and reliability of the different delivery approaches, should also come into play when making this decision. These issues were beyond the scope of this review but should be examined in future work to help inform this decision making.

We identified that initial engagement was high across the few studies that reported this outcome. We would also have liked to examine long-term engagement and the effect of this on intervention success, but the highly inconsistent and sporadic reporting of long-term engagement within included studies made synthesizing this information impossible in this review. There is no consensus regarding the definition of engagement in the context of eHealth and how to measure it, as many studies of behavior change interventions do not report at all on their participant engagement characteristics, yet report on attrition [65-69]. People could report absolute frequencies of contact, and where relevant, duration of contact. However, a recent new concept of effective engagement outlines that for each different intervention delivered, and for each individual person receiving the intervention, their requirement to engage and over what period of time will be different to achieve the intended outcome [70]. This concept indicates that *absolute* measures of engagement in eHealth interventions may not tell the full story as to whether the individual has engaged to a degree that is optimal for them. Perski et al [71] proposed that engagement with digital behavior change interventions occurs through specific direct and indirect mechanisms of action. Constructs of content, delivery, context, target behavior, population, and setting were proposed as important mechanisms to understand and report on when understanding engagement.

Our findings contrast with a previous review that looked at telephone interventions on physical activity and dietary behavior change in the population without cancer [72], in that the magnitude of effect appears to be lower than previously pooled analysis of behavior change outcomes using telephone interventions (effect size 0.60; 95% CI 0.24 to 1.19). It is plausible that the benefit of eHealth interventions in cancer populations may be diminished compared with populations without cancer because of concurrent treatment side effects and the impact that a cancer diagnosis can have on the capacity of patients to absorb new information [73]. A systematic review [74] of the effectiveness of mHealth technology use in behavior change interventions demonstrated mixed results. Studies reporting benefits described small effect sizes that were retained only in the short term. An older review [10] of interventions promoting health behavior change via the internet also examined the mode of delivery on efficacy and included 3 categories: automated functions, communicative functions, and supplementary modes. This review found that interventions were more effective in eliciting behavior change when there was more extensive use of underlying behavior change theory, more techniques of behavior change used in the delivery of the intervention, and additional methods of interacting with participants (combined techniques). This review found that between 2 and 17 behavior change techniques were used in each study, and a range of theoretical underpinnings was employed. This variability may also have contributed to the variability seen in the results of this review. These previous reviews were limited by not examining the potential mechanisms of action leading to the behavior change or the moderating effect of how the interventions were delivered.

We acknowledge a number of limitations on how we decided to undertake this review. Many of the studies included in this review used self-reported outcome measures with many failing to also include objective measures to corroborate data. We also combined both objective and subjective measures of behavior in the same meta-analysis. The small number of studies and high heterogeneity of data precluded a multiple metaregression to assess the interaction between behavior change and intervention delivery mode on health outcomes. Response bias, including social desirability bias, occurs frequently where self-reported outcome measures are used in research [75]. Ecological fallacy could also explain why we did not find a relationship between the amount of behavior change and the amount of change in quality of life. This could have been brought about by the examination of study level comparisons of analytic approaches and relationships rather than at the person level. We need to be able to trace the health behavior change data to the individual participants to understand the impact of the intervention.

Many of the studies reported multiple outcome measures, each with differing results. In this review, we chose to use data from the primary outcome measures (where reported) as the outcome of choice for behavior change and health outcomes. Where this was not reported, we decided to choose the outcome that we believed to be the most relevant outcome to the intervention examined in that particular study. We chose not to conduct metaregression analyses comparing synchronous and asynchronous eHealth approaches for outcomes where only one study was available in either of these subgroups. Such analyses would arguably have had limited generalizability and were at high risk of committing a type II statistical error. We chose to include studies within this synthesis regardless of the type of cancer involved or of the behavior change theories and techniques employed. Each theory and technique may have a different effect on the outcomes we examined and may be a source of confounding for our comparisons of synchronous and asynchronous eHealth approaches. Similarly, the type of patient population may have a moderating effect on the efficacy of synchronous and asynchronous eHealth interventions. Ideally, a comparison of synchronous and asynchronous eHealth approaches would be undertaken in the same populations, using the same behavior change theories and techniques. However, no such studies were identified in this review.

Effective health behavior change has been ascribed to rely on the use of behavior change theory. In all, 83% (n=20/24) of studies included in this review reported on the use of behavior change theory. It was beyond the scope of this review to detail how behavior change theory was used to develop the various interventions used in the included studies. This is an area of great interest and could be pursued in future reviews seeking to identify how interventions are constructed, using the methods of Michie and Prestwich [76].

There are a range of studies that looked at synchronous and asynchronous methods in health behavior change interventions in cancer patients and/or survivors, but there is a complete absence of RCTs that compared the differences in the delivery methods within one specific trial. This is where the gap in evidence lies. This systematic review highlights the need for further 3-arm studies comparing both synchronous and asynchronous interventions compared with a standard care group. There should also be an economic evaluation to determine which is also the most cost-effective intervention. Robust reporting of engagement, not only at the initiation of the trial but throughout, is also essential to gain a greater understanding of the complexity of participant engagement in study efficacy and how to replicate this in future implementation of eHealth interventions. Trials of this nature will enable the determination of the most successful method of delivery in terms of effectiveness, acceptability, user engagement, cost-effectiveness, successful behavior change, and ultimately translation into health outcomes.

Conclusions

This systematic review and meta-analysis provides evidence that behavior change interventions delivered via eHealth, particularly on physical activity and diet modification delivered to cancer patients or survivors, show benefit. There is insufficient evidence to determine whether the specific delivery mode of eHealth (synchronous, asynchronous, or combined) modulates this effectiveness. Three-arm RCTs comparing asynchronous and synchronous delivery modes with a control with robust engagement reporting are required to determine the most successful delivery method for promoting behavior change and ultimately favorable health outcomes.

Acknowledgments

The authors wish to thank Monash Health for providing support for this project.

Authors' Contributions

KF was the lead reviewer and drafted the manuscript. MS was the secondary reviewer for full-text screening. MS and KH were secondary reviewers for data extraction. KH was the secondary reviewer for the risk of bias assessment. KF completed the data analysis and TH reviewed for accuracy. All authors (MS, KH, TH, and DC) provided feedback on the draft manuscript. All authors (MS, KH, TH, and DC) reviewed the final manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred reporting items for systematic reviews and meta-analysis checklist.
[DOC File , 65 KB - jmir_v22i6e16112_app1.doc]

Multimedia Appendix 2
SCOPUS search strategy.
[PNG File , 252 KB - jmir_v22i6e16112_app2.png]

Multimedia Appendix 3
Descriptive summary.
[DOCX File , 35 KB - jmir_v22i6e16112_app3.docx]

Multimedia Appendix 4
Behavior change theories and techniques.
[DOCX File , 27 KB - jmir_v22i6e16112_app4.docx]

References

1. World Health Organization. 2016. Global Diffusion Of eHealth: Making Universal Health Coverage Achievable URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf;jsessionid=B0BB29A9D5D9F63C603D92C66E0852E6?sequence=1 [accessed 2019-10-06]
2. World Health Organization. 2011. mHealth: New Horizons for Health through Mobile Technology URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2019-10-06]
3. International Telecommunication Union (ITU). 2019. Statistics URL: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2019-02-14]
4. Organisation for Economic Co-Operation and Development. 2018. Internet Access URL: https://data.oecd.org/ict/internet-access.htm [accessed 2019-09-20]
5. Budman SH, Portnoy D, Villapiano AJ. How to get technological innovation used in behavioral health care: build it and they still might not come. Psychotherapy 2003;40(1-2):45-54. [doi: 10.1037/0033-3204.40.1-2.45]
6. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]

7. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]
8. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
9. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
10. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
11. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
12. International Telecommunication Union (ICT). 2017. Measuring Digital Development: Facts and Figures 2019 URL: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [accessed 2019-02-14]
13. Halwas N, Griebel L, Huebner J. eHealth literacy, internet and ehealth service usage: a survey among cancer patients and their relatives. J Cancer Res Clin Oncol 2017 Nov;143(11):2291-2299. [doi: 10.1007/s00432-017-2475-6] [Medline: 28699035]
14. Hsu J, Huang J, Kinsman J, Fireman B, Miller R, Selby J, et al. Use of e-health services between 1999 and 2002: a growing digital divide. J Am Med Inform Assoc 2005;12(2):164-171 [FREE Full text] [doi: 10.1197/jamia.M1672] [Medline: 15561786]
15. World Health Organisation. 2018. Latest Global Cancer Data: Cancer Burden Rises to 18.1 Million New Cases and 9.6 Million Cancer Deaths in 2018 URL: https://www.who.int/cancer/PRGlobocanFinal.pdf [accessed 2020-01-18]
16. Aziz NM, Rowland JH. Trends and advances in cancer survivorship research: challenge and opportunity. Semin Radiat Oncol 2003 Jul;13(3):248-266. [doi: 10.1016/S1053-4296(03)00024-9] [Medline: 12903014]
17. Petrick JL, Reeve BB, Kucharska-Newton AM, Foraker RE, Platz EA, Stearns SC, et al. Functional status declines among cancer survivors: trajectory and contributing factors. J Geriatr Oncol 2014 Oct 1;5(4):359-367 [FREE Full text] [doi: 10.1016/j.jgo.2014.06.002] [Medline: 24981125]
18. Okwuosa TM, Anzevino S, Rao R. Cardiovascular disease in cancer survivors. Postgrad Med J 2017 Feb;93(1096):82-90. [doi: 10.1136/postgradmedj-2016-134417] [Medline: 28123076]
19. Kushi LH, Doyle C, McCullough M, Rock CL, Demark-Wahnefried W, Bandera EV, American Cancer Society 2010 Nutrition and Physical Activity Guidelines Advisory Committee. American cancer society guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin 2012;62(1):30-67 [FREE Full text] [doi: 10.3322/caac.20140] [Medline: 22237782]
20. Doyle C, Kushi LH, Byers T, Courneya KS, Demark-Wahnefried W, Grant B, 2006 Nutrition, Physical Activity and Cancer Survivorship Advisory Committee, American Cancer Society. Nutrition and physical activity during and after cancer treatment: an American cancer society guide for informed choices. CA Cancer J Clin 2006;56(6):323-353 [FREE Full text] [doi: 10.3322/canjclin.56.6.323] [Medline: 17135691]
21. Cancer Council Australia. 2013. Position Statement - Benefits of Health Diet and Physical Activity for Cancer Survivors URL: https://wiki.cancer.org.au/policy/Position_statement_-_Benefits_of_healthy_diet_and_physical_activity_for_cancer_survivors [accessed 2020-01-18]
22. Kiss N, Baguley BJ, Ball K, Daly RM, Fraser SF, Granger CL, et al. Technology-supported self-guided nutrition and physical activity interventions for adults with cancer: systematic review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e12281 [FREE Full text] [doi: 10.2196/12281] [Medline: 30747720]
23. Goode AD, Lawler SP, Brakenridge CL, Reeves MM, Eakin EG. Telephone, print, and web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. J Cancer Surviv 2015 Dec;9(4):660-682. [doi: 10.1007/s11764-015-0442-2] [Medline: 25757733]
24. Kern J. Evaluation of teleconsultation systems. Int J Med Inform 2006;75(3-4):330-334. [doi: 10.1016/j.ijmedinf.2005.08.001] [Medline: 16143564]
25. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May 1;4(3):666-684 [FREE Full text] [doi: 10.1177/193229681000400323] [Medline: 20513335]
26. Wootton R. Twenty years of telemedicine in chronic disease management-an evidence synthesis. J Telemed Telecare 2012 Jun;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
27. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google is here to stay but health care professionals are still valued: an analysis of health care consumers' internet navigation support preferences. J Med Internet Res 2017 Jun 14;19(6):e210 [FREE Full text] [doi: 10.2196/jmir.7489] [Medline: 28615156]

28. Graffigna G, Barello S, Bonanomi A, Riva G. Factors affecting patients' online health information-seeking behaviours: the role of the patient health engagement (PHE) model. Patient Educ Couns 2017 Oct;100(10):1918-1927. [doi: 10.1016/j.pec.2017.05.033] [Medline: 28583722]
29. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Training. 2019 Jul. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 URL: https://training.cochrane.org/handbook [accessed 2020-05-29]
30. Kirkpatrick D, Kirkpatrick J. Evaluating Training Programs: The Four Levels. Third Edition. Oakland, CA: Berrett-Koehler Publishers; 2005.
31. Higgins J, Altman D, Sterne J. Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, USA: The Cochrane Collaboration; 2011.
32. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, USA: Academic Press; 1977.
33. Lee MK, Yun YH, Park H, Lee ES, Jung KH, Noh D. A web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
34. Higgins J, Li T, Deeks J. Choosing effect measures and computing estimates of effect. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, USA: Wiley; 2019.
35. Eakin EG, Lawler SP, Winkler EA, Hayes SC. A randomized trial of a telephone-delivered exercise intervention for non-urban dwelling women newly diagnosed with breast cancer: exercise for health. Ann Behav Med 2012 Apr;43(2):229-238. [doi: 10.1007/s12160-011-9324-7] [Medline: 22109352]
36. Hawkes AL, Chambers SK, Pakenham KI, Patrao TA, Baade PD, Lynch BM, et al. Effects of a telephone-delivered multiple health behavior change intervention (CanChange) on health and behavioral outcomes in survivors of colorectal cancer: a randomized controlled trial. J Clin Oncol 2013 Jun 20;31(18):2313-2321. [doi: 10.1200/JCO.2012.45.5873] [Medline: 23690410]
37. Ligibel JA, Meyerhardt J, Pierce JP, Najita J, Shockro L, Campbell N, et al. Impact of a telephone-based physical activity intervention upon exercise behaviors and fitness in cancer survivors enrolled in a cooperative group setting. Breast Cancer Res Treat 2012 Feb;132(1):205-213 [FREE Full text] [doi: 10.1007/s10549-011-1882-7] [Medline: 22113257]
38. Freeman MF, Tukey JW. Transformations related to the angular and the square root. Ann Math Statist 1950 Dec;21(4):607-611. [doi: 10.1214/aoms/1177729756]
39. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
40. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 1927 Jun;22(158):209. [doi: 10.2307/2276774]
41. Bowen DJ, Burke W, Hay JL, Meischke H, Harris JN. Effects of web-based intervention on risk reduction behaviors in melanoma survivors. J Cancer Surviv 2015 Jun;9(2):279-286 [FREE Full text] [doi: 10.1007/s11764-014-0412-0] [Medline: 25425205]
42. Djuric Z, Ellsworth JS, Weldon AL, Ren J, Richardson CR, Resnicow K, et al. A diet and exercise intervention during chemotherapy for breast cancer. Open Obes J 2011;3:87-97 [FREE Full text] [doi: 10.2174/1876823701103010087] [Medline: 22238561]
43. Mayer DK, Landucci G, Awoyinka L, Atwood AK, Carmack CL, Demark-Wahnefried W, et al. SurvivorCHESS to increase physical activity in colon cancer survivors: can we get them moving? J Cancer Surviv 2018 Feb;12(1):82-94 [FREE Full text] [doi: 10.1007/s11764-017-0647-7] [Medline: 28994035]
44. Morey MC, Snyder DC, Sloane R, Cohen HJ, Peterson B, Hartman TJ, et al. Effects of home-based diet and exercise on functional outcomes among older, overweight long-term cancer survivors: RENEW: a randomized controlled trial. J Am Med Assoc 2009 May 13;301(18):1883-1891 [FREE Full text] [doi: 10.1001/jama.2009.643] [Medline: 19436015]
45. Bantum EO, Albright CL, White KK, Berenberg JL, Layi G, Ritter PL, et al. Surviving and thriving with cancer using a web-based health behavior change intervention: randomized controlled trial. J Med Internet Res 2014 Feb 24;16(2):e54 [FREE Full text] [doi: 10.2196/jmir.3020] [Medline: 24566820]
46. Parsons JK, Newman VA, Mohler JL, Pierce JP, Flatt S, Marshall J. Dietary modification in patients with prostate cancer on active surveillance: a randomized, multicentre feasibility study. BJU Int 2008 May;101(10):1227-1231 [FREE Full text] [doi: 10.1111/j.1464-410X.2007.07365.x] [Medline: 18218061]
47. Parsons JK, Pierce JP, Natarajan L, Newman VA, Barbier L, Mohler J, et al. A randomized pilot trial of dietary modification for the chemoprevention of noninvasive bladder cancer: the dietary intervention in bladder cancer study. Cancer Prev Res (Phila) 2013 Sep;6(9):971-978 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-13-0050] [Medline: 23867158]
48. Pinto B, Dunsiger S, Stein K. Does a peer-led exercise intervention affect sedentary behavior among breast cancer survivors? Psychooncology 2017 Nov;26(11):1907-1913 [FREE Full text] [doi: 10.1002/pon.4255] [Medline: 27531024]
49. Porter LS, Gao X, Lyna P, Kraus W, Olsen M, Patterson E, et al. Pilot randomized trial of a couple-based physical activity videoconference intervention for sedentary cancer survivors. Health Psychol 2018 Sep;37(9):861-865. [doi: 10.1037/hea0000608] [Medline: 30138021]

50. Rabin C, Dunsiger S, Ness KK, Marcus BH. Internet-based physical activity intervention targeting young adult cancer survivors. J Adolesc Young Adult Oncol 2011 Dec;1(4):188-194 [FREE Full text] [doi: 10.1089/jayao.2011.0040] [Medline: 23610737]
51. Hatchett A, Hallam JS, Ford MA. Evaluation of a social cognitive theory-based email intervention designed to influence the physical activity of survivors of breast cancer. Psychooncology 2013 Apr;22(4):829-836. [doi: 10.1002/pon.3082] [Medline: 22573338]
52. Emmons KM, Puleo E, Sprunck-Harrild K, Ford J, Ostroff JS, Hodgson D, et al. Partnership for health-2, a web-based versus print smoking cessation intervention for childhood and young adult cancer survivors: randomized comparative effectiveness study. J Med Internet Res 2013 Nov 5;15(11):e218 [FREE Full text] [doi: 10.2196/jmir.2533] [Medline: 24195867]
53. Dieng M, Butow PN, Costa DS, Morton RL, Menzies SW, Mireskandari S, et al. Psychoeducational intervention to reduce fear of cancer recurrence in people at high risk of developing another primary melanoma: results of a randomized controlled trial. J Clin Oncol 2016 Dec 20;34(36):4405-4414. [doi: 10.1200/JCO.2016.68.2278] [Medline: 27998215]
54. Kim SH, Shin MS, Lee HS, Lee ES, Ro JS, Kang HS, et al. Randomized pilot test of a simultaneous stage-matched exercise and diet intervention for breast cancer survivors. Oncol Nurs Forum 2011 Mar;38(2):E97-106. [doi: 10.1188/11.ONF.E97-E106] [Medline: 21356647]
55. Yun YH, Lee KS, Kim Y, Park SY, Lee ES, Noh D, et al. Web-based tailored education program for disease-free cancer survivors with cancer-related fatigue: a randomized controlled trial. J Clin Oncol 2012 Apr 20;30(12):1296-1303. [doi: 10.1200/JCO.2011.37.2979] [Medline: 22412149]
56. Kanera IM, Bolman CA, Willems RA, Mesters I, Lechner L. Lifestyle-related effects of the web-based Kanker Nazorg Wijzer (cancer aftercare guide) intervention for cancer survivors: a randomized controlled trial. J Cancer Surviv 2016 Oct;10(5):883-897 [FREE Full text] [doi: 10.1007/s11764-016-0535-6] [Medline: 26984534]
57. Ormel HL, van der Schoot GG, Westerink ND, Sluiter WJ, Gietema JA, Walenkamp AM. Self-monitoring physical activity with a smartphone application in cancer patients: a randomized feasibility study (SMART-trial). Support Care Cancer 2018 Nov;26(11):3915-3923 [FREE Full text] [doi: 10.1007/s00520-018-4263-5] [Medline: 29785635]
58. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Oct 30;15(1):106 [FREE Full text] [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
59. Forbes CC, Blanchard CM, Mummery WK, Courneya KS. Feasibility and preliminary efficacy of an online intervention to increase physical activity in Nova Scotian cancer survivors: a randomized controlled trial. JMIR Cancer 2015 Nov 23;1(2):e12 [FREE Full text] [doi: 10.2196/cancer.4586] [Medline: 28410166]
60. Villaron C, Cury F, Eisinger F, Cappiello M, Marqueste T. Telehealth applied to physical activity during cancer treatment: a feasibility, acceptability, and randomized pilot study. Support Care Cancer 2018 Oct;26(10):3413-3421. [doi: 10.1007/s00520-018-4191-4] [Medline: 29675546]
61. Pinto BM, Stein K, Dunsiger S. Peers promoting physical activity among breast cancer survivors: a randomized controlled trial. Health Psychol 2015 May;34(5):463-472 [FREE Full text] [doi: 10.1037/hea0000120] [Medline: 25110844]
62. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
63. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
64. Cox M, Basen-Engquist K, Carmack CL, Blalock J, Li Y, Murray J, et al. Comparison of internet and telephone interventions for weight loss among cancer survivors: randomized controlled trial and feasibility study. JMIR Cancer 2017 Sep 27;3(2):e16 [FREE Full text] [doi: 10.2196/cancer.7166] [Medline: 28954716]
65. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
66. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
67. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
68. Short C, Rebar A, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Eur Psychol 2015;17(1):32 [FREE Full text]
69. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010 Feb 10;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]

70. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
71. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
72. Eakin EG, Lawler SP, Vandelanotte C, Owen N. Telephone interventions for physical activity and dietary behavior change: a systematic review. Am J Prev Med 2007 May;32(5):419-434. [doi: 10.1016/j.amepre.2007.01.004] [Medline: 17478269]
73. Cornwell P, Dicks B, Fleming J, Haines TP, Olson S. Care and support needs of patients and carers early post-discharge following treatment for non-malignant brain tumour: establishing a new reality. Support Care Cancer 2012 Oct;20(10):2595-2610. [doi: 10.1007/s00520-012-1383-1] [Medline: 22286363]
74. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
75. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res 2011 Oct;2(4):320-332 [FREE Full text] [doi: 10.1504/IJBHR.2011.043414] [Medline: 25383095]
76. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]

Abbreviations

eHealth: electronic health
mHealth: mobile health
PRISMA: preferred reporting items for systematic reviews and meta-analysis
RCTs: randomized controlled trials
SMD: standardized mean difference

Edited by G Eysenbach; submitted 03.09.19; peer-reviewed by M Ponum, D Chao, A Roberts, T Freeman; comments to author 22.10.19; revised version received 19.11.19; accepted 09.04.20; published 23.06.20.

Please cite as:
Furness K, Sarkies MN, Huggins CE, Croagh D, Haines TP
Impact of the Method of Delivering Electronic Health Behavior Change Interventions in Survivors of Cancer on Engagement, Health Behaviors, and Health Outcomes: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(6):e16112
URL: http://www.jmir.org/2020/6/e16112/
doi:10.2196/16112
PMID:32574147

Commentary

How to Fight an Infodemic: The Four Pillars of Infodemic Management

Gunther Eysenbach[1], MD, MPH, FACMI

[1] JMIR Publications, Toronto, ON, Canada

Corresponding Author:
Gunther Eysenbach, MD, MPH, FACMI
JMIR Publications
130 Queens Quay East, Suite 1100
Toronto, ON,
Canada
Phone: 1 416 583 2040
Email: editor@jmir.org

Related Article:

Comment on: http://www.jmir.org/2020/6/e19659/

Abstract

In this issue of the *Journal of Medical Internet Research*, the World Health Organization (WHO) is presenting a framework for managing the coronavirus disease (COVID-19) infodemic. Infodemiology is now acknowledged by public health organizations and the WHO as an important emerging scientific field and critical area of practice during a pandemic. From the perspective of being the first “infodemiologist” who originally coined the term almost two decades ago, I am positing four pillars of infodemic management: (1) information monitoring (infoveillance); (2) building eHealth Literacy and science literacy capacity; (3) encouraging knowledge refinement and quality improvement processes such as fact checking and peer-review; and (4) accurate and timely knowledge translation, minimizing distorting factors such as political or commercial influences. In the current COVID-19 pandemic, the United Nations has advocated that facts and science should be promoted and that these constitute the antidote to the current infodemic. This is in stark contrast to the realities of infodemic mismanagement and misguided upstream filtering, where social media platforms such as Twitter have advertising policies that sideline science organizations and science publishers, treating peer-reviewed science as “inappropriate content.”

(J Med Internet Res 2020;22(6):e21820) doi:10.2196/21820

KEYWORDS

infodemiology; infodemic; COVID-19; infoveillance; pandemic; epidemics; emergency management; public health

The World Health Organization Declares an Infodemic and Crowdsources a Framework

In this issue of the *Journal of Medical Internet Research*, a high-profile group of authors associated with the World Health Organization (WHO) have published a paper entitled “Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation” [1]. In the paper, the authors collected and organized global ideas to fight the current coronavirus disease (COVID-19) infodemic declared by the WHO on February 15, 2020. Impressively, this consultation meeting was entirely conducted online, and, as noted by the authors, turned out to be one of the largest meetings ever convened by the WHO.

In my capacity as the Editor of the *Journal of Medical Internet Research*, I had the honor of attending the meeting as an invited panelist to talk about the responsibility and potential approaches for scholarly publishers to combat the infodemic. A summary of my presentation and a project proposal for scientists, editors, and science communicators will be the subject of a separate, forthcoming editorial.

I also attended the meeting as one of the early researchers in this field. In fact, I coined the terms *infodemiology* [2,3], *infodemic*, and *infoveillance* [4] over the course of the past 20 years. It is not without pride that I witnessed how this line of research is now formally acknowledged by public health organizations and the WHO as a novel, emerging scientific field and critical area of practice during a pandemic [1]. It should also be noted that this journal has been a pioneer in encouraging and disseminating this line of research and has provided a central forum for researchers to discuss and publish their infodemiology work in a high-impact journal [5-7].

A Brief History of Infodemiology

Although the term *infodemiology* was coined in 2002 [2], concerns over infodemics or outbreaks of misinformation are almost as old as the World Wide Web itself. A widely cited paper published in 1997 in the *BMJ* drew attention to the, now seemingly trivial, finding that medical information found on the internet is not always reliable [8]. This paper sparked an avalanche of publications describing and analyzing the quality of medical information on different topics, which I, together with my colleagues, synthesized in a comprehensive systematic review published in *JAMA* in 2002 [9]. At the time, I was also leading major EU-funded projects (MedCERTAIN and MedCIRCLE) attempting to label health information on the internet with machine-readable metadata [10,11], an ambitious project with the goal to create something akin to a digital “immune system” for misinformation [12], with distributed descriptive and evaluative metadata as machine-processable “antibodies.” The metadata (evaluative or descriptive, supplied by third parties or the providers themselves) are perhaps conceptually comparable to contemporary efforts by Facebook and Twitter to fact-check information and label social media posts that are problematic [13], although these tags are not necessarily machine processable and cannot be harvested by third-party applications, which was the underlying idea of the MedCERTAIN/MedCIRCLE projects [10,11]. The goal was to create a global infrastructure for such machine-processable annotations that would allow humans and machines to attain a more complete picture about what other people and institutions have to say or think about a certain information provider or piece of information. The vast amount of information on the web, the dynamic nature of the web, and questions on the scalability of this approach were obvious limitations, but, perhaps it is time, for the sake of future infodemic preparedness, to revisit some of these “semantic web” ideas articulated over 20 years ago and to combine them with today’s powerful artificial intelligence tools, because, given the advances in natural language processing, many of these metadata labels could now be generated automatically. My former project partner Dan Brickley is now working with Google and is running schema.org [14], which organizes community vocabularies to enable such applications.

I coined the term “infodemiology” in 2002 in a short guest editorial in the *American Journal of Medicine* [2], defining “infodemiology” as a “new emerging research discipline and methodology” comprising the “the study of the determinants and distribution of health information and misinformation—which may be useful in guiding health professionals and patients to quality health information on the Internet.” Equipped with the awareness that “quality of health information” as well as “misinformation” is often hard to define (as quality is in the eye of the beholder, and a “fact” in medicine requires more than one patient or study), I framed infodemiology as a method to “identify areas where there is a knowledge translation gap between best evidence (what some experts know) and practice (what most people do or believe)” [2]. While this early work focused on information *supply* (what is published on the internet), I added, in 2006, analysis of information *demand* (search queries) to the concept, realizing that harvesting what people are searching for on the internet could inform areas of public health such as surveillance. I illustrated this with demonstrating the predictive power of Internet searches to predict flu outbreaks [3,4]; an idea that inspired Google Flu Trends [15]. With the emergence of Twitter, more “social listening” infoveillance studies became possible, and H1N1 (SwineFlue) became the first pandemic where this approach could be demonstrated; my graduate student Cynthia Chew and I, analyzed the content of pandemic tweets and determined, among other interesting findings, a prevalence of misinformation of 4.5% [16].

In a tweet posted on April 14, 2020, Secretary General of the United Nations (UN) António Guterres announced a UN communications response initiative countering the infodemic with facts and evidence [17]; however, what we have learned in 20 years of infodemiology research is that the quality of health information is an elusive concept, as in medicine, the truth is not always easy to determine, especially in a rapidly evolving situation.

While certain technical quality criteria, readibility scores, and the compliance with ethical quality criteria (such as the presence of disclosure of who owns the site and conflicts of interests, all aspects that are important to determine the source credibility) are relatively easy to measure, the concepts of *accuracy*, *facts*, and *truth* usually require the presence of evidence-based guidelines or systematic reviews as a gold standard to determine what works and what does not. In a rapidly evolving situation such as the COVID-19 pandemic, some of the problems are the rapid rate of new scientific information published and the inability of researchers, policy makers, journalists, and ordinary citizens to keep up with quickly changing facts. In other words, the current pandemic is partly a challenge to filter (in real time) the sheer quantity of information published on a daily basis. The founder of the website Retraction Watch, Ivan Oransky, stresses that science is a conversation [18]. Even a publication of a clinical study is not the last word and studies may be contradicted or proven wrong. In the early phases of a pandemic, "facts" are perhaps more accurately referred to as "BETs" (best evidence at the time). Facts are sparse and recommendations based on BETs are subject to change. The COVID-19 pandemic has illustrated this with examples such as mask-wearing recommendations, use of certain drugs such as hydroxychloroquine, and social distancing or school opening guidelines. The public health and medical evidence also needs to be integrated with economic and political considerations and may be subject to cultural variations and influences. Thus, the proposal to fight the infodemic by spreading “facts” is easier said than done when it is not clear what the exact facts are.

The Information “Cake” Model

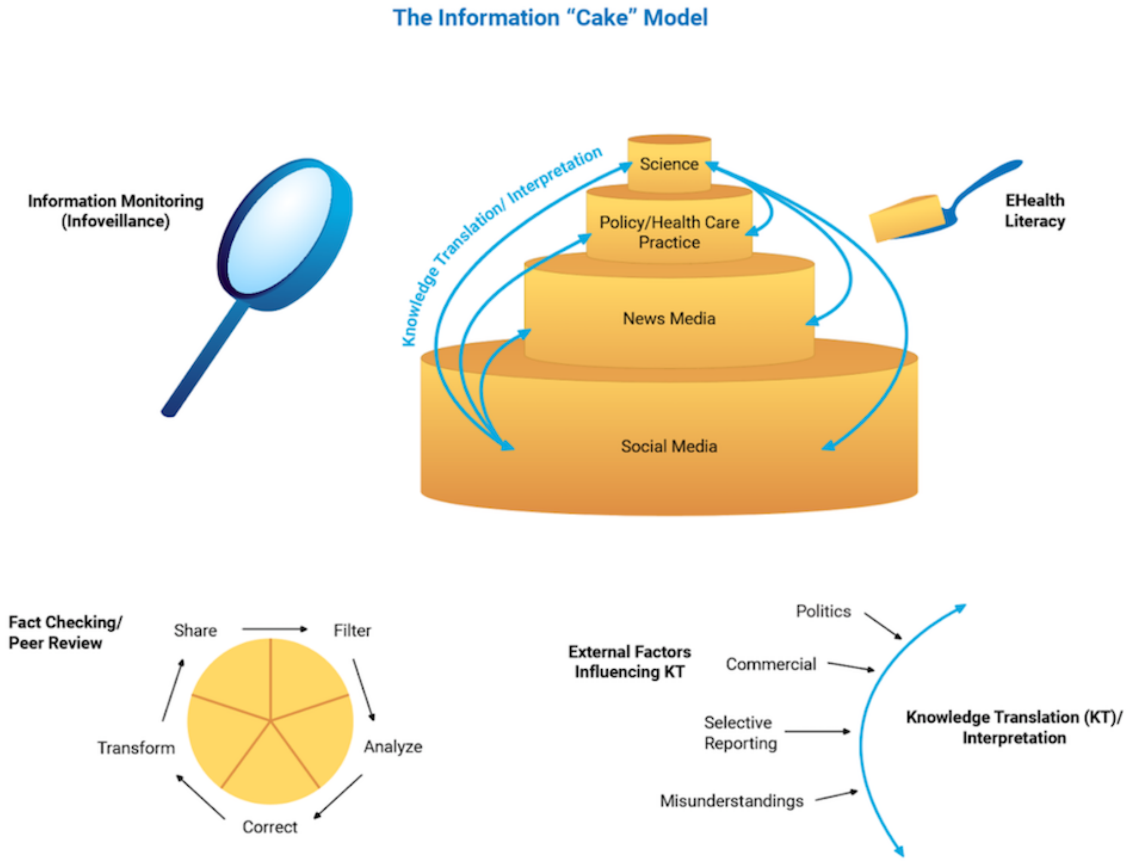
The following model is not the framework presented in the WHO paper in this journal. It is an "expert opinion" (to the degree as somebody - and probably the first - who has "Infodemiologist" in his LinkedIn profile can be considered an "expert"). It supplements the WHO framework by providing a first broad roadmap on how to fight an infodemic.

The current infodemic is a crisis to distill the sheer quantity of information, which is occurring on four levels: (1) science, (2) policy and practice, (3) news media, and (4) social media.

The wedding cake model (Figure 1) illustrates these four levels as layers. The size of the layers is proportional to the amount of information generated by these four groups of actors. The model also shows some information flows and knowledge translation activities that take place between these different levels. Science is the smallest layer of the wedding cake in terms of the amount of information, and it is depicted at the top of the information wedding cake, which represents rigorous and selective information production cycles. Clearly, misinformation can be found here as well, perhaps measured by the number of retractions, which, as of June 2020, stands at less than two dozen retracted articles [19], but this number is certain to increase. With over 26,000 COVID-19 articles indexed in Pubmed, this represents less than 0.1% of the published research, even though there may be a higher rate within the segment of unreviewed preprints, some of which may never see the light of journal publication, which may be another metric for the prevalence of scientific misinformation (somebody please do a study on this and submit it to the *Journal of Medical Internet Research*!). The main problem is not so much the prevalence of misinformation in the science layer, but the challenge of translating this information into actionable recommendations and conveying conclusions for different audiences and stakeholders in other layers, illustrated by the knowledge translation arrows in Figure 1.

Social media is depicted as the largest and last segment of the wedding cake, representing the vast amount of nearly unfiltered and uncontrolled information generated or amplified by the public. Information in social media is, of course, also generated by science organizations, policy makers, health care organizations, and journalists.

Figure 1. The Information "Cake" Model. The four pillars of infodemic management are information monitoring (infoveillance; top left); building eHealth Literacy and science literacy (top right); encouraging knowledge refinement and quality improvement processes for information providers, such as fact checking and peer review (bottom left); and Knowledge Translation, meaning to translate knowledge from one layer to another, while minimizing distorting factors (bottom right). eHealth: electronic health; KT: knowledge translation.



First Pillar: Facilitate Accurate Knowledge Translation

Between the levels, knowledge translation processes need to take place to translate the information from one audience to another, and these knowledge translation processes are, perhaps, the main mechanisms where information becomes misinformation, as the interpretation of "facts" is subject to multiple potentially influencing factors such as politics, commercial interests, selective reporting, and misunderstandings. These knowledge translation processes take place between all four layers of the cake, for example, between public health policy recommendations and news media (to keep the figure simple, not all possible knowledge translation arrows between the different layers are shown in Figure 1). *The first* *pillar of infodemic management is to support, facilitate, and strengthen accurate knowledge translation*. In the WHO paper, authors hint at that when they write “knowledge should be translated into actionable behaviour-change messages presented in ways that are understood by and accessible to all individuals” [1], but this is of course just one example for a KT problem, while there are also other KT challenges, e.g. between science and policy makers. Realizing that knowledge translation is subject to political, commercial, or other influences that distort the scientific message, the influencing factors should be minimized or, if present, at least clearly disclosed and called out.

Second Pillar: Knowledge Refinement, Filtering, and Fact-Checking

The FACTS model [20], depicted as an insert in Figure 1, is an example for a fact checking workflow, and reminds us that within each layer, there are knowledge refinement processes such as fact checking and quality assurance mechanisms at play, which are sometimes visible and sometimes invisible to the end user. For example, on the science level, the process of peer reviewing and publishing scholarly work is a method to constantly filter, refine, and improve the information generated by previous scholars. This also means that within each layer, there is a continuum ranging from raw, tentative, and possibly problematic information to highly refined and trustworthy information. *The second pillar of infodemic management is to encourage, facilitate, and strengthen knowledge refinement and filtering processes on each level, to accelerate internal quality improvement processes.* Within each layer, information in different stages of this "refinement" process can be found (for example, in the science layer, unreviewed preprints, laying right next to peer-reviewed scholarly communications); thus, clearly labelling the provenance of the information at the different knowledge production stages is equally as important as facilitating and accelerating them..

Third pillar: Build eHealth Literacy

The cake-serving utensil illustrates that in the age of the internet and openness, the end user is able to (but not always equipped) to consume information from any level, in any refinement stage, making eHealth literacy an essential skill in a networked world. eHealth literacy is defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem." [21,22]. In the information age (which is now perhaps shifting to the infodemic age!), the user carries a significant part of the responsibility to select and downstream-filter trustworthy health information. For example, nothing stops a user from tapping into the vast array of unreviewed preprints published in preprint servers such as medRxiv, but interpreting and contextualizing the information found here requires significant eHealth literacy (which includes science literacy) skills. Thus, *the third pillar of infodemic management is to enhance the capacity of all stakeholders to build eHealth literacy, to select and assess health and science information found on the different layers of the information cake*. This aspect is notably underdeveloped in the WHO paper’s taxonomy but can be seen as part of WHO’s “identify evidence” category.

Fourth Pillar: Monitoring, Infodemiology, Infoveillance, and Social Listening

The fourth pillar of infodemic management is continuous monitoring and analysis of data and information exchange patterns on the internet, a field I have called *infodemiology and infoveillance* [4]. My idea was that similar to surveillance in pandemics we want to be able to detect outbreaks of misinformation, rumors, falsehoods, to counter them with facts or other interventions. Infoveillance requires generating metrics on information supply on the internet, including its quality (for example incidence of anti-vaccination tweets), as well as information demand metrics, such as search queries or questions posed on social media or other web 2.0 platforms. In Figure 1, infoveillance is illustrated as a lens magnifying the information exchange patterns within different communities and for different subtopics.

Conclusions

Poorly executed and uncoordinated infodemic management may lead to unintended consequences such as the sidelining and suppression of science in favor of political and commercial interests.

Such an unintended consequence is demonstrated, for example, by the poorly thought-through advertising policy of Twitter [23], which only allows governments and selected news media, but not science organization or science publishers, to amplify messages. Under this “inappropriate content” policy of Twitter, only the following kinds of tweets are allowed to be amplified and promoted [23]:

- *Public service announcements related to COVID-19 from governments and supranational entities (for example, World Health Organization) as well as trusted partners approved by the Public Policy team*
- *News related to COVID-19 from media publishers who have been exempted under the Political Ads policy.*

Notably missing from Twitters' exemption list are science organizations and science publishers. Is this indicative of a sidelining of science in favor of politics, or just an oversight? JMIR Publications (as science publisher) ran into this problem first-hand when we were prevented by Twitter from disseminating COVID-19 peer-reviewed research, promoting our virtual COVID-19 preprint journal clubs, etc. Therefore, if the UN declares an infodemic and promotes science and “facts” as the antidote, then the suppression of science as “inappropriate content” by private social media platforms should be an alarming sign that indicates that there is ample room of improvement in how the current infodemic is managed and coordinated by different stakeholders.

The current COVID-19 pandemic is a 9/11 for public health, but also an opportunity to develop and formalize tools and approaches for future infodemic management. It is also an opportunity to re-engineer certain knowledge refinement processes such as scholarly publishing and peer-review (stayed tuned to what we are doing with JMIRx.org). Much as improvements in information flows between government agencies post-9/11 helped to prevent another major act of terrorism in the United States, improved and preventive infodemic prevention and management can mitigate the next infodemic, which we will face as soon as a vaccine is available. The price for freedom of speech and improved information technology is an increased susceptibility to infodemics. We are entering the age of infodemics.

Conflicts of Interest

The author is founder and president of JMIR Publications.

References

1. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
2. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [Medline: 12517369]
3. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
4. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
5. Mavragani A. Infodemiology and Infoveillance: Scoping Review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
6. Barros JM, Duggan J, Rebholz-Schuhmann D. The Application of Internet-Based Sources for Public Health Surveillance (Infoveillance): Systematic Review. J Med Internet Res 2020 Mar 13;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]
7. E-collection 'Infodemiology and Infoveillance'. Journal of Medical Internet Research. URL: https://www.jmir.org/themes/69 [accessed 2020-06-26]
8. Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. BMJ 1997 Jun 28;314(7098):1875-1879 [FREE Full text] [doi: 10.1136/bmj.314.7098.1875] [Medline: 9224132]
9. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
10. Eysenbach G, Diepgen TL. Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ 1998 Nov 28;317(7171):1496-1500 [FREE Full text] [doi: 10.1136/bmj.317.7171.1496] [Medline: 9831581]
11. Eysenbach G, Yihune G, Lampe K, Cross P, Brickley D. Quality management, certification and rating of health information on the Net with MedCERTAIN: using a medPICS/RDF/XML metadata structure for implementing eHealth ethics and creating trust globally. J Med Internet Res 2000;2(2 Suppl):2E1 [FREE Full text] [Medline: 11720933]
12. Coiera E. Information epidemics, economics, and immunity on the internet. We still know so little about the effect of information on public health. BMJ 1998 Nov 28;317(7171):1469-1470 [FREE Full text] [doi: 10.1136/bmj.317.7171.1469] [Medline: 9831570]
13. Roth Y. Updating our approach to misleading information. Twitter Blog. 2020 May 11. URL: https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html [accessed 2020-06-26]
14. Schema.org. URL: https://schema.org/ [accessed 2020-06-26]
15. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
16. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
17. António Guterres @antonioguterres. Twitter. 2020 Apr 14. URL: https://twitter.com/antonioguterres/status/1250095790959267841?s=20 [accessed 2020-06-26]
18. Yan W. Coronavirus Tests Science?s Need for Speed Limits. The New York Times. URL: https://www.nytimes.com/2020/04/14/science/coronavirus-disinformation.html [accessed 2020-06-26]
19. Retracted coronavirus (COVID-19) papers. Retraction Watch. URL: https://retractionwatch.com/retracted-coronavirus-covid-19-papers/ [accessed 2020-06-26]

20. Coulombe C. April 2020. FACTS Framework for Fighting Misinformation and Disinformation. Forthcoming COVID19. URL: http://www.covid19facts.ca/
21. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
22. E-collection 'eHealth Literacy / Digital Literacy'. Journal of Medical Internet Research. URL: https://www.jmir.org/themes/70 [accessed 2020-06-26]
23. Inappropriate content. Twitter Business. URL: https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/inappropriate-content.html [accessed 2020-06-26]

Abbreviations

BET: Best evidence at the time
COVID-19: coronavirus disease
eHealth: electronic health
UN: United Nations
WHO: World Health Organization

Submitted 25.06.20; this is a non–peer-reviewed article;accepted 26.06.20; published 29.06.20.

Please cite as:
Eysenbach G
How to Fight an Infodemic: The Four Pillars of Infodemic Management
J Med Internet Res 2020;22(6):e21820
URL: http://www.jmir.org/2020/6/e21820/
doi:10.2196/21820
PMID:32589589

Viewpoint

Rethinking the Use of Mobile Apps for Dietary Assessment in Medical Research

Wael Khazen[1*], PhD; Jean-François Jeanne[2], MSc; Laëtitia Demaretz[2]; Florent Schäfer[2*], MSc; Guy Fagherazzi[3,4*], MSc, PhD

[1]Kelly OCG, Paris, France
[2]Innovation Science and Nutrition, Danone Nutricia Research, Palaiseau, France
[3]Digital Epidemiology Hub, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg
[4]Center of Research in Epidemiology and Population Health, UMR 1018 Inserm, Institut Gustave Roussy, Paris-Sud Paris-Saclay University, Villejuif, France
[*]these authors contributed equally

Corresponding Author:
Florent Schäfer, MSc
Innovation Science and Nutrition
Danone Nutricia Research
RD 128 Avenue de la Vauve
Palaiseau, 91767
France
Phone: 33 622256853
Email: florent.schafer@danone.com

Abstract

Food intake and usual dietary intake are among the key determinants of health to be assessed in medical research and important confounding factors to be accounted for in clinical studies. Although various methods are available for gathering dietary data, those based on innovative technologies are particularly promising. With combined cost-effectiveness and ease of use, it is safe to assume that mobile technologies can now optimize tracking of eating occasions and dietary behaviors. Yet, choosing a dietary assessment tool that meets research objectives and data quality standards remains challenging. In this paper, we describe the purposes of collecting dietary data in medical research and outline the main considerations for using mobile dietary assessment tools based on participant and researcher expectations.

(J Med Internet Res 2020;22(6):e15619) doi:10.2196/15619

KEYWORDS

diet; dietary assessment; epidemiology; clinical research; mobile diet app; academic apps; consumer-grade apps

Diet Assessment in Clinical Studies: Purposes and Limitations

In which situations should diet be assessed in clinical research? Which considerations should drive the selection of a dietary assessment tool (DAT)? This paper addresses these two questions, emphasizing that one should consider not only the research objectives and study setting, but also the opportunities and limitations of available methods for acquiring and analyzing dietary data.

In epidemiological and clinical observational studies, dietary parameters are adequately considered when information on usual dietary intake is needed at a population level [1] or when diet-disease relationships are being explored [2]. In these settings, evaluating dietary intake over time has proved efficient for investigating diet-disease associations at the population level, especially in chronic conditions [2]. Tracking participant lifestyle was needed to understand the role of unhealthy food and diet in the development of overweight, obesity, type 2 diabetes [3], cardiovascular diseases [3,4], and cancer [5]. However, despite the availability of scientifically validated methods to assess dietary intake and usual intake, their use in establishing diet-disease relationships remains controversial [6,7].

In interventional settings, dietary parameters need to be considered for efficacy assessments or to control confounding effects, irrespective of the nature (dietary, drug, or other) of the intervention [8]. However, background diet is, to our knowledge, not systematically assessed in drug development studies despite evidence of the presence of food-drug interactions.

Food-drug interaction is also an important factor to consider during clinical development, as some chemical compounds in foods can potentially affect the pharmacokinetic, pharmacodynamic, or metabolic pathways of some drugs [9]. For example, it is known that leafy green vegetables such as kale and spinach, which contain high levels of vitamin K, can reduce the effectiveness of some oral anticoagulants, such as warfarin [10]. In parallel, we must not neglect how the side effects of medicinal products can influence appetite [11].

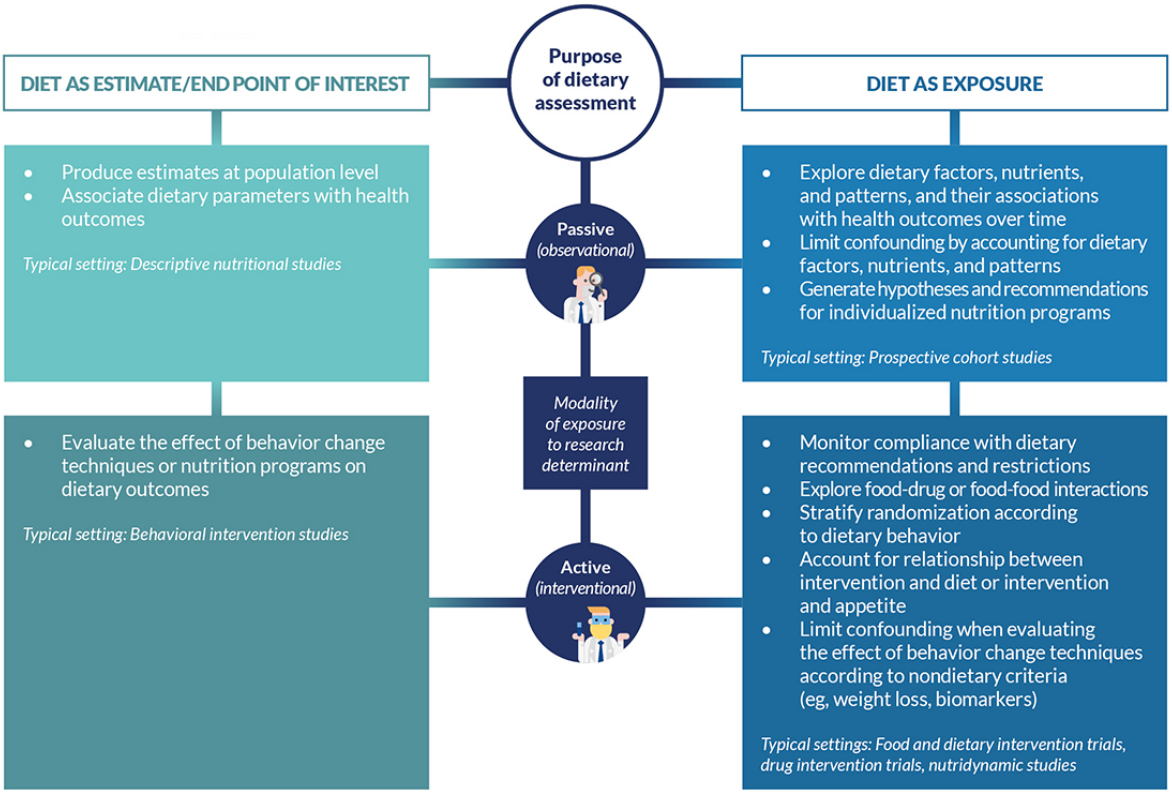
Choosing the Optimal DAT

Selecting a DAT adapted to a research question is usually a matter of compromise. Current methods to assess diet are available in various formats (paper or electronic), to obtain data on dietary intake (prospectively or retrospectively) or usual intake [12]. In a comprehensive resource guide issued by the Food and Agriculture Organization of the United Nations [13] and as explained by other authors [14], various available DATs can be distinguished based on technology (including data acquisition):

- Conventional methods for dietary data collection, which include food records, food frequency questionnaires, 24-hour dietary recalls and diet history, which are widely used in research, with known strengths and limitations
- DATs based on innovative technologies, which are divided into 5 categories according to the acquisition method: personal digital assistant, image-assisted methods (ie, digital cameras), mobile-based technology, interactive computer and web-based technologies, and scan and sensor-based technologies. It must be noted that DATs based on innovative technologies can be digitized versions of the above-mentioned conventional methods.

Collecting dietary data is, however, not always needed for primary study analyses, as one may only be interested in monitoring the background diet or compliance with dietary restrictions while focusing on many other clinical and nonclinical parameters (Figure 1).

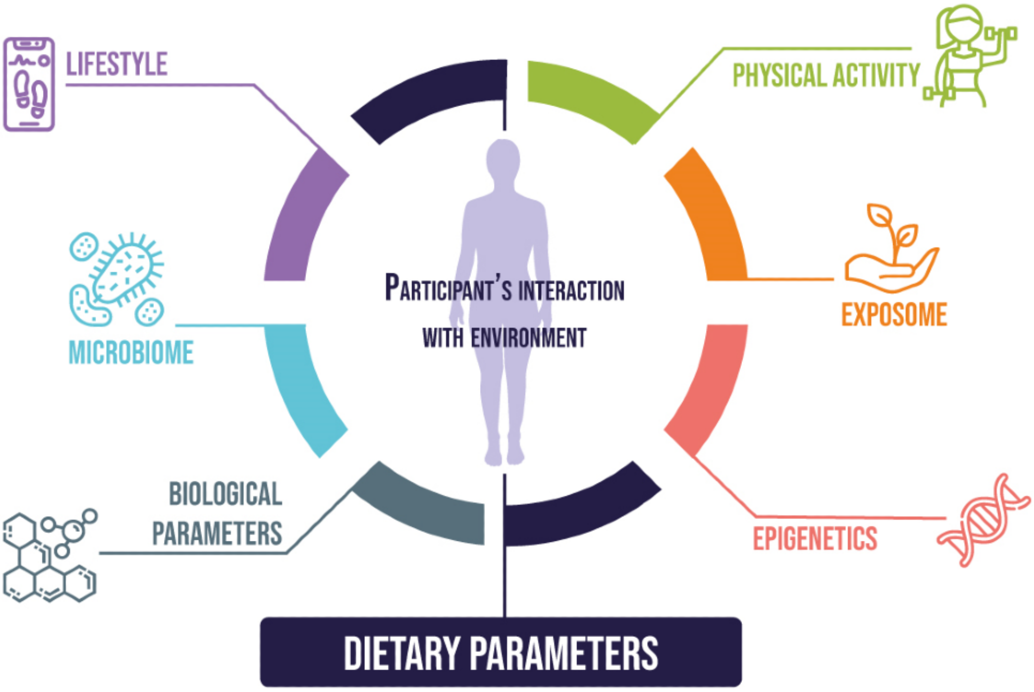
Figure 1. Relevant frameworks for using dietary assessment tools in human studies.



Because collecting dietary information usually requires substantial involvement from clinical study participants, selecting an appropriate DAT represents a trade-off between participant burden and the volume or accuracy of the collected information. Compromises may include collecting dietary intake data at specific time points rather than tracking each eating occasion, and choosing a reasonable frequency for assessing food intake throughout the research studies.

However, limiting the dietary assessment to habitual food intake (evaluated using food frequency questionnaires) is not relevant for every study, especially when associations between dietary and biological parameters are investigated, as outlined in Figure 2.

Figure 2. Parameters that can be measured or accounted for along with dietary parameters in medical research.



Instead of evaluating habitual food intake, daily tracking of each eating occasion may be needed. One example is the daily variation of gut microbiota composition, known to be related to food choice [15], requiring, when investigated, exhaustive and daily tracking of a participant's dietary intake. Handling temporality, therefore, represents a major challenge when selecting a relevant DAT, in case there is important intrasubject variability of parameters over time. We believe that a focus is needed for identifying or developing solutions that can effectively track participants' background diet throughout their participation in research studies (observational or interventional), irrespective of the intervention (drug, food, or dietary) while limiting the associated participant burden. DATs based on new technologies offer opportunities to facilitate traditional food intake measures, and mobile apps have demonstrated being adequate for both participants and researchers. For participants, mobile apps represent an acquisition method that is acceptable, accessible, ubiquitous, and one that can be used by participants alone [16-18]. For the study sponsors or researchers, these apps offer other opportunities in terms of cost-effectiveness, reducing time between the collection and reporting of dietary data, and improving data quality with an accurate, comprehensive, and relevant food composition data set [14].

Mobile apps, therefore, represent a good compromise for researchers (investigators and sponsors) when the collection of dietary data is needed.

The Need to Explore the True Potential of Mobile Apps as DATs

When navigating the astronomical number of mobile diet apps (approximately 30,000 diet and physical activity mobile apps are currently available on Google Play and the App Store), selecting a DAT that fits with research purposes remains a key challenge. This selection should be based on multiple parameters, including the suitability of user experience and interface, the reliability of the food composition database (FCD), the relevance of the data acquisition method (identification and quantification of foods), and the quality of nutritional outputs.

To compare these mobile diet apps in terms of reliability, quality, and effectiveness, we cannot consider the number of downloads as a reliable evaluation parameter, as the most frequently downloaded apps are not necessarily the most usable or efficient in medical research [19]. Neither can we assess these tools based on user ratings (ie, stars), which are known to be biased [20]. Currently, no official standards exist for evaluating health-related mobile apps, making the evaluation of the strengths and weaknesses of these apps complex for researchers when choosing a relevant tool [21].

High-quality reviews and meta-analyses are few, and there is limited comparability between the evaluation methods used. Some of the reviews evaluate functionality, information quality, esthetics, and engagement [22]; others evaluate the accountability, usability, scientific coverage, and technology-enhanced features based on the opinions of nutrition experts who had tested different apps for 5 days [23]. In parallel, other experts have developed their own scales for a qualitative assessment of mobile health apps, including the following:

- Mobile App Rating Scale, which focuses on 4 evaluation parameters by an expert panel, including engagement, functionality, aesthetics, and quality of the information provided by the app [24]
- App Quality Evaluation, which evaluates the educational quality and technical functionality of nutrition apps by nutrition professionals and app users, based on several parameters, including app function, purpose, behavior change potential, support of knowledge acquisition, and skill development [25]

To our knowledge, only 3 studies have compared the nutritional values provided by different mobile diet apps based on quantitative parameters. One research team evaluated the accuracy of 7 diet apps in providing nutrient values. In their study, the nutrient labels of 100 food products available in the Netherlands were compared to the nutrient values provided in the nutrition facts in each app. The researchers concluded that the accuracy of nutrient values varied enormously between these apps, and energy was the most reliable value [26]. Another research team compared the nutrient intake calculations of 5 consumer-grade mobile diet apps to the Nutrition Data System for Research dietary analysis program, concluding that most of the nutritional values (except energy and some macronutrients) provided in these apps are generally underestimated, and that some apps are more accurate than others [27].

More recently, another team evaluated energy, macronutrient, and available micronutrient values provided by 5 popular nutrition apps against a UK reference method (Dietplan6 [28]). The authors concluded that the values of energy, carbohydrates, saturated fats, total fat, and fiber reported by these apps (except for 1 app) were rather reliable, while the values of protein, sodium, and micronutrients were inconsistent and less reliable [29].

Academic and Consumer-Grade Mobile Diet Apps

Two Main Categories of Mobile Diet Apps

Even though mobile diet apps are emerging tools, 2 main categories can be distinguished based on the affiliation of the developers and their objectives:

- Academic apps, developed by experts in nutrition or dietetics to provide a reliable, scientifically validated tool; these are mainly developed for research purposes. Known examples include Technology Assisted Dietary Assessment [30], DietCam [31], and My Meal Mate [32].
- Consumer-grade apps that are service-oriented, typically developed by private entities specialized in digital development [33]. These apps are intended to be used by the lay public, and their purpose is mainly commercial. Popular examples include MyFitnessPal, FatSecret, and Lifesum. Weight management is usually the main feature of this second category of apps.

The main strengths and limitations of each category, from the participant and researcher perspective, are outlined in Figure 3.

Academic Apps: A Focus on Scientific Validation

From a researcher's point of view, academic apps offer more advantages than consumer-grade apps, as they are usually developed with identifiable scientific input and compared with a standard method for validation purposes. Typical examples of academic apps include Electronic Dietary Intake Assessment (e-DIA) [34] and My Meal Mate [32]. Academic apps are primarily focused on research and do not aim to be popular. For most of these apps, the number of users is unknown (as they are usually not referenced in Google Play or the App Store), but we assume it is lower than the number of users of consumer-grade apps, which aim to be visible in the stores.

Another main advantage of academic apps is the absence of any constraints in terms of advertising and commercial incentives for users. We also believe that these apps are safer in terms of privacy and confidentiality than those designed for consumer use.

A recent systematic assessment of technology-based DATs [35] compared the features of research and consumer tools. The authors concluded that features facilitating data entry (including voice, digital images, and bar code scan) were more frequently available in “consumer” apps, such as DietCam or My Meal Mate. This work, however, did not consider the “consumer-grade” apps that are discussed in this paper. Other examples of academic apps or apps developed with identifiable academic input that include such features are MyFoodRepo (which includes an image recognition algorithm) [36] and FODMAP App for people with irritable bowel syndrome (FoodMaestro bar code scan, a commercial app) [37]. Most academic apps are not designed for universal use as they are validated for specific homogenous populations, which prevents their use in an international clinical study. Similarly, another important difference between these 2 categories is the number of available languages, as most academic apps are usually available in 1 or 2 languages at most (myfood24 [38] is available in English, German, and Danish, and developers are currently working on the release of an Arabic version).

The question of the scientific validation of academic apps remains a matter of debate. In general, academic apps are evaluated using another self-report instrument thought to capture diet as a reference, such as 24-hour dietary recalls. The choice of the reference method can be controversial, as stated by the National Cancer Institute, which considers that data collected using this method contain errors, including intake-related bias [39]. It should be noted that the e-DIA app was validated by comparison to 24-hour dietary recall as a reference [34], and the same reference method was used to validate My Meal Mate app [32]. Results of validation studies must, therefore, be interpreted with caution when scientific validation is relative and not absolute. These validation methods are less reliable than using gold-standard reference measures, such as biomarker recovery, or methods to capture true intake without systematic errors, such as direct observation and feeding studies [39].

Consumer-Grade Apps: Volume of FCD Data and User Experience

Given the lucrative purpose of consumer-grade apps, an app must be attractive to appeal to a maximum number of users. Even though this environment is competitive, some of these apps are very popular and are downloaded tens of millions of times around the world (such as MyFitnessPal, FatSecret, and Lose It!). For the end user, the apps designed as a service usually offer an appealing user interface and experience: their design is attractive, which facilitates the identification of a meal and may reduce the time required to enter data compared to some academic apps. Another major difference between these 2 categories is the number of languages available (for example, MyFitnessPal is available in 19 languages [26]). The use of a very rich and international FCD is usually a key differentiator among these apps. However, the quality of the FCDs, regularly pointed at as a major issue with consumer-grade apps, is discussed later in this article.

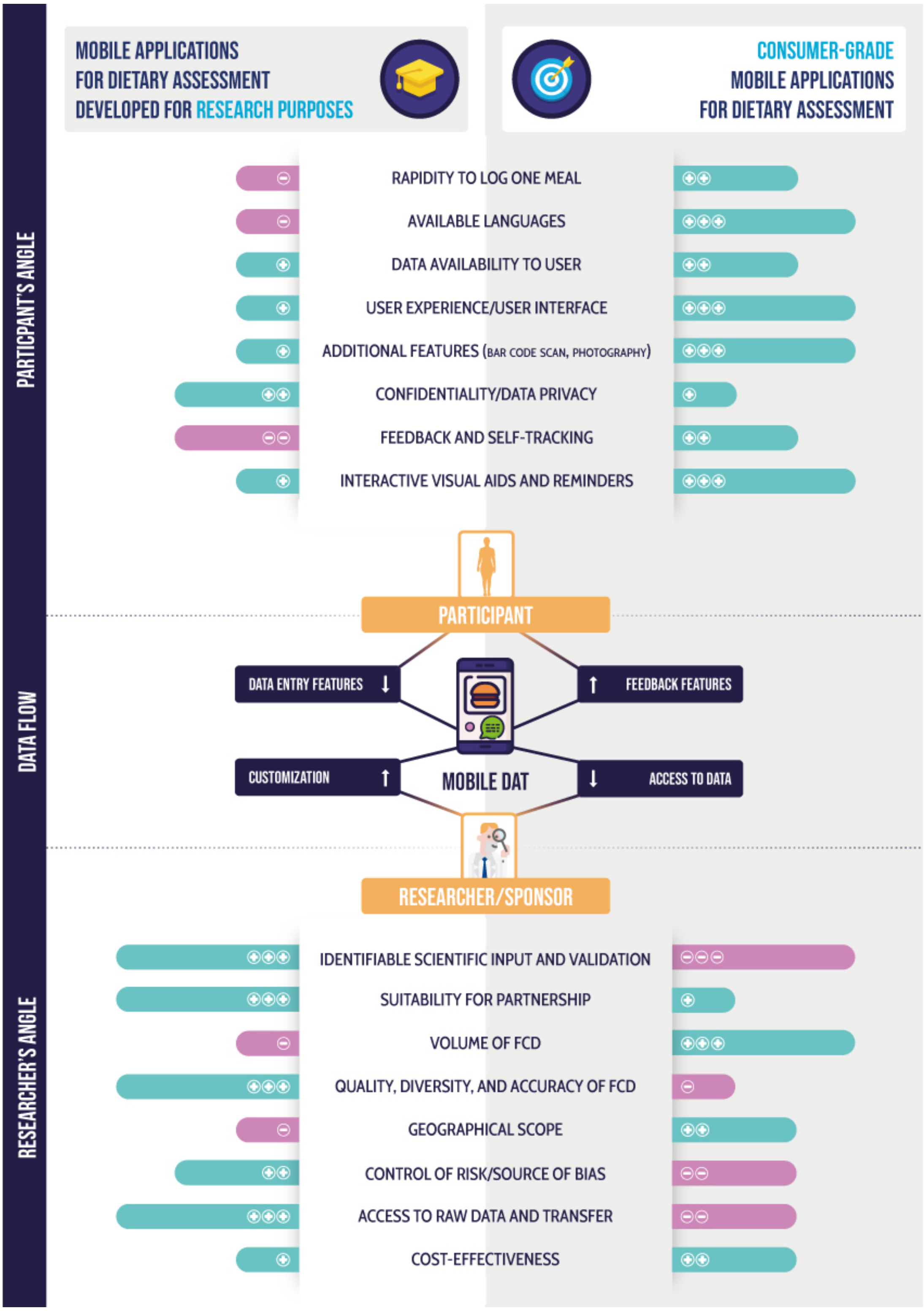
The usability of consumer-grade apps is now being explored but is difficult to assess. A recent review [40] concluded that the features available in consumer apps vary greatly, and there are additional variations between Android and iOS versions. The key features of these apps usually include interactive visual aids and reminders, which are available in basic free versions (even though some apps require subscriptions) to help users achieve their goals. Thus, to further engage users in improving these apps, end-users can add new food products that are not available in the FCD, including energy and nutrient values, or to correct inaccurate information [41]. This may be beneficial for FCD enrichment, but it may also be a major and dangerous source of error in data entry. The addition of new food products and associated nutrient values by lay users can lead to input errors, especially when the developers of these apps do not control the quality and integrity of the data sets, and this is one of the most important negative features of consumer-grade apps.

On the other hand, the use of these consumer-grade apps in medical research has certain limitations and risks, especially with regard to data privacy and confidentiality. The lack of documentation regarding the privacy of user data may represent a major challenge for the legal and ethical integrity of personal and food data [42]. A recent example is the data breach of 150 million accounts of MyFitnessPal [43], which is considered the world's no. 1 app in the field of consumer-grade diet-tracking apps. Another major limitation of these apps is the strong incentive to purchase a paid version through advertisements and recurring notifications: these may cause annoyance to the users of a basic free version and represent a bias in research when too many recommendations are provided as they may interfere with evaluation of the parameters.

We can safely assume that most consumer-grade apps have been developed without rigorous scientific validation and do not systematically involve nutrition experts during development. A recent review of 28 905 relevant health-oriented mobile apps showed that only 17 apps (0.05%) were developed with identifiable professional input [44]; other analyses showed that only 0.8% (3/393) of weight management apps have been scientifically evaluated [45].

Consumer-grade apps may, however, offer some benefits to researchers, such as FCD data volume, as it differs hugely between these 2 categories of apps. In academic apps, the approximate number of food items in the FCD is tens of thousands at most (8500 for DietCam [31] and 40,000 for My Meal Mate [35]), while this number is much higher in several consumer-grade apps (eg, 18 million foods claimed by MyFitnessPal developers [46]).

Figure 3. Key aspects of mobile dietary assessment tools (DATs) developed for research purposes and consumer use from the participant and researcher perspectives. FCD: food composition database.



Room for Improvement of Consumer-Grade Apps

If the volume of FCD data is a key differentiator for these 2 categories of apps, the quality, diversity, and accuracy of FCDs is widely in favor of academic apps. Indeed, the quality of FCDs of consumer-grade apps remains questionable, as explained by several authors [26,27,29] who have concluded that these apps lack a reliable FCD; provide a limited set of nutrients to users; are too focused on energy and macronutrient intake, with a lot of missing data; and underestimate macronutrient and micronutrient values.

Several developers assume that some food information in their database is inaccurate, and some of them, such as MyFitnessPal [41] or Lose It! [47], invite users to correct these errors using crowdsourcing techniques.

Academic apps are regarded as more reliable in their FCD, based on objective measurements and scientific studies [48,49]. For researchers and sponsors, other benefits are in favor of academic apps, including the suitability for partnership, as the developers of these types of apps are usually more open to share information about the development and validation of these tools. Considering that these apps are primarily designed for research, the access to and transfer of raw data are facilitated for researchers with dedicated access. For apps focused on consumers, the access to data is mainly provided to users, especially those who subscribe to optional features. In addition, the absence of feedback features in academic apps reduces the risk of bias in the tracking of eating occasions. However, it should be noted that the geographic scope of consumer-grade apps, which are developed and available worldwide, is far more significant than that of academic apps, which are usually limited to one or two countries.

Discussion

In summary, we can surmise that researchers will usually favor using academic mobile apps, while end users may be more attracted to the features offered by consumer-grade mobile apps. Based on our experience and the cited opinions of experts in the field of nutrition, we understand that currently, there is no one-size-fits-all solution for tracking dietary intake that can be used in every type of medical research, as each available tool has its limitations [35,50] due to language availability, validation in specific populations, and the reliability of FCDs. The ideal option would be to develop a mobile app for tracking the diet that has the strengths of both types of apps: an efficient, user-friendly interface and experience, coupled with an FCD that is as rich as it is reliable. In the absence of such a tool, or in developing one soon, we must select the most appropriate one for each type of study according to its objectives. The landmark DIET@NET (DIETary Assessment Tool NETwork) suggests considering 5 steps before choosing the appropriate type of DAT, and the first step is, of course, to define the research question well [50].

If the daily use of consumer apps by millions of people worldwide can generate big data on food consumption, this information can be used by researchers (while complying with data privacy requirements) to study health, diet, and dietary behavior parameters. This can be done by considering cultural aspects of populations at different levels (country, household, or individual) while integrating potential key cofactors such as physical activity level, sleep, weight, or location data, which can often be collected simultaneously. It has also been highlighted that the most frequently downloaded consumer-grade apps lack features relating to emotions, even though emotions are known to be associated with diet [40]. Using these diet apps as a reliable food-tracking tool in clinical trials or research, in general, is a major challenge, considering their limitations; despite the popularity of these apps, the lack of evidence and scientific validation of their use remains the major issue [51]. Little information is available on the quality of these mobile apps, other than what can be gathered from users' ratings published on mobile app stores. The lack of reliability of the data these apps provide, especially in the FCD, is a real barrier for researchers who want to utilize the user data. It should also be noted that the legal and ethical concerns related to user data for consumer-grade mobile health apps are well documented in the literature [52-54]. A recent review showed that most mobile diet apps do not provide terms of use or privacy policy documents [42].

The use of consumer diet apps in studies should be done with caution: given the nature and suboptimal quality of their FCDs, their use should be limited to the tracking of some macroelements such as energy intake. Energy intake is considered the most consistently and accurately value reported by these apps compared to other macro- and micronutrients [23,26], which are generally underestimated [27,29,55]. Recent studies have also shown that different consumer apps, such as Samsung Health, MyFitnessPal, and FatSecret, provide an acceptable estimate of energy, carbohydrate, and fat intake [29], and that MyFitnessPal offers a good relative validity for energy and fiber tracking [55]. Many dieticians (in the United Kingdom, New Zealand, and Australia) use nutrition apps in their practice [56]. However, for many experts in the field of nutrition, mobile apps designed for consumer awareness are an unreliable source of nutritional values due to the crowdsourcing nature of the FCD. This issue is leading to a lack of interest in using these new tools for nutritional and clinical studies.

Perspectives

Our opinion is that the potential of mobile apps designed for consumer use has not been sufficiently evaluated, especially in clinical settings. Despite the limitations that have been outlined above, we believe that the data collected by these apps represent a source of information we cannot ignore, as it may be sufficient in some epidemiological or clinical studies, when the assessment of diet is not the primary objective of a research study. Potential uses of these tools include segmenting participant populations into subgroups according to usual dietary intake or associating specific meal patterns with biological parameters during a short time frame. However, in our view, further work is needed to assess this potential. We should also consider the potential of other features that are now included in several mobile apps, such as automated, image-based recognition. This feature, while promising, is only in an early stage of development, and there is still room for improvement until it can be used regularly by clinical study participants. However, when relevant, analytical methods offered by artificial intelligence may soon be used to obtain both qualitative (ie, composition of meals) and quantitative (ie, portion sizes) outputs. We believe that such features may one day be the bridge that accommodates participants and researchers, filling the gap between academic and consumer-grade apps once they are available in both types of solutions.

Acknowledgments

WK, FS, and GF authored the manuscript and created the figures. LD and JFJ reviewed the manuscript. All authors validated the final version of the manuscript and agreed to be accountable for the accuracy of the work. The authors thank Mrs Livia Recchia for the graphic design of the figures and Mrs Debjani Roy for her thorough proofreading of the manuscript.

Conflicts of Interest

WK is employed by a contractor of Danone Nutricia Research. JFJ, LD, and FS are employed by Danone Nutricia Research. GF has received consulting fees from Danone Nutricia Research.

References

1. Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. Adv Nutr 2015 Jan;6(1):5-18 [FREE Full text] [doi: 10.3945/an.114.007492] [Medline: 25593140]
2. Darnton-Hill I, Nishida C, James W. A life course approach to diet, nutrition and the prevention of chronic diseases. Public Health Nutr 2007 Jan 02;7(1a):101-121. [doi: 10.1079/phn2003584] [Medline: 14972056]
3. Abdullah A, Peeters A, de Courten M, Stoelwinder J. The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies. Diabetes Res Clin Pract 2010 Sep;89(3):309-319. [doi: 10.1016/j.diabres.2010.04.012] [Medline: 20493574]
4. Strazzullo P, D'Elia L, Cairella G, Garbagnati F, Cappuccio FP, Scalfi L. Excess body weight and incidence of stroke: meta-analysis of prospective studies with 2 million participants. Stroke 2010 May;41(5):e418-e426. [doi: 10.1161/STROKEAHA.109.576967] [Medline: 20299666]
5. Fiolet T, Srour B, Sellem L, Kesse-Guyot E, Allès B, Méjean C, et al. Consumption of ultra-processed foods and cancer risk: results from NutriNet-Santé prospective cohort. BMJ 2018 Dec 14;360:k322 [FREE Full text] [doi: 10.1136/bmj.k322] [Medline: 29444771]
6. Ioannidis JPA. Reforming Nutritional Epidemiologic Research-Reply. JAMA 2019 Dec 22;321(3):310. [doi: 10.1001/jama.2018.18676] [Medline: 30667499]
7. Archer E, Lavie CJ, Hill JO. The Failure to Measure Dietary Intake Engendered a Fictional Discourse on Diet-Disease Relations. Front Nutr 2018;5:105 [FREE Full text] [doi: 10.3389/fnut.2018.00105] [Medline: 30483510]
8. Schäfer F, Jeanne J. Evaluating the effects of food on health in a world of evolving operational challenges. Contemp Clin Trials Commun 2018 Dec;12:51-54 [FREE Full text] [doi: 10.1016/j.conctc.2018.09.001] [Medline: 30259002]
9. Fujita K. Food-drug interactions via human cytochrome P450 3A (CYP3A). Drug Metabol Drug Interact 2004;20(4):195-217. [doi: 10.1515/dmdi.2004.20.4.195] [Medline: 15663291]
10. Holmes MV, Hunt BJ, Shearer MJ. The role of dietary vitamin K in the management of oral vitamin K antagonists. Blood Rev 2012 Jan;26(1):1-14. [doi: 10.1016/j.blre.2011.07.002] [Medline: 21914559]
11. White R. Drugs and nutrition: how side effects can influence nutritional intake. Proc Nutr Soc 2010 Nov;69(4):558-564. [doi: 10.1017/S0029665110001989] [Medline: 20678294]
12. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. F1000Res 2017;6:926 [FREE Full text] [doi: 10.12688/f1000research.10703.1] [Medline: 28690835]
13. Food and Agriculture Organization of the United Nations. 2017. Guidelines on Assessing Biodiverse Foods in Dietary Intake Surveys URL: http://www.fao.org/3/a-i6717e.pdf [accessed 2020-03-19] [WebCite Cache ID 78XVFPpYt]
14. Amoutzopoulos B, Steer T, Roberts C, Cade JE, Boushey CJ, Collins CE, et al. Traditional methods new technologies - dilemmas for dietary assessment in large-scale nutrition surveys and studies: a report following an international panel discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisbane, 3 September 2015. J Nutr Sci 2018;7:e11 [FREE Full text] [doi: 10.1017/jns.2018.4] [Medline: 29686860]
15. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, Personalized Microbiome Class Students, et al. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. Cell Host Microbe 2019 Jun 12;25(6):789-802.e5. [doi: 10.1016/j.chom.2019.05.005] [Medline: 31194939]
16. Lieffers JRL, Valaitis RF, George T, Wilson M, Macdonald J, Hanning RM. A Qualitative Evaluation of the eaTracker Mobile App. Nutrients 2018 Oct 09;10(10):e1462 [FREE Full text] [doi: 10.3390/nu10101462] [Medline: 30304766]
17. Bucher DTS, Carrard I, Farina E, Kruseman M. Usability and Acceptability of a Mobile Application to Track Food Intake in a Research Context. J Acad Nutr Diet 2016 Sep;116(9):A61. [doi: 10.1016/j.jand.2016.06.213]
18. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
19. Zaidan S, Roehrer E. Popular Mobile Phone Apps for Diet and Weight Loss: A Content Analysis. JMIR Mhealth Uhealth 2016 Jul 11;4(3):e80 [FREE Full text] [doi: 10.2196/mhealth.5406] [Medline: 27400806]
20. Girardello A. AppAware: which mobile applications are hot? In: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services. 2010 Sep Presented at: MobileHCI '10; September 2010; Lisbon p. 431-434 URL: https://dl.acm.org/doi/10.1145/1851600.1851698 [doi: 10.1145/1851600.1851698]

21. Xu W, Liu Y. mHealthApps: A Repository and Database of Mobile Health Apps. JMIR Mhealth Uhealth 2015;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
22. Bardus M, van Beurden S, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
23. Chen J, Cade JE, Allman-Farinelli M. The Most Popular Smartphone Apps for Weight Loss: A Quality Assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. DiFilippo KN, Huang W, Chapman-Novakofski KM. A New Tool for Nutrition App Quality Evaluation (AQEL): Development, Validation, and Reliability Testing. JMIR Mhealth Uhealth 2017 Oct 27;5(10):e163 [FREE Full text] [doi: 10.2196/mhealth.7441] [Medline: 29079554]
26. Maringer M, Wisse-Voorwinden N, Veer PV, Geelen A. Food identification by barcode scanning in the Netherlands: a quality assessment of labelled food product databases underlying popular nutrition applications. Public Health Nutr 2019 May;22(7):1215-1222. [doi: 10.1017/S136898001800157X] [Medline: 29962361]
27. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. Public Health Nutr 2018 Dec;21(8):1495-1502. [doi: 10.1017/S1368980018000393] [Medline: 29534771]
28. McCance RA. In: Roe MA, editor. McCance and Widdowson's The Composition of Foods, 6th edition, summary edition. Cambridge: Royal Society of Chemistry; 2002.
29. Fallaize R, Zenun Franco R, Pasang J, Hwang F, Lovegrove JA. Popular Nutrition-Related Mobile Apps: An Agreement Assessment Against a UK Reference Method. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e9838 [FREE Full text] [doi: 10.2196/mhealth.9838] [Medline: 30785409]
30. Ahmad Z, Kerr DA, Bosch M, Boushey CJ, Delp EJ, Khanna N, et al. A Mobile Food Record For Integrated Dietary Assessment. MADiMa16 (2016) 2016 Oct;2016:53-62 [FREE Full text] [doi: 10.1145/2986035.2986038] [Medline: 28691119]
31. Kong F, Tan J. DietCam: Automatic dietary assessment with mobile camera phones. Pervasive Mob Comput 2012 Feb;8(1):147-163 [FREE Full text] [doi: 10.1016/j.pmcj.2011.07.003]
32. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2013 Feb 14;109(3):539-546. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
33. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83. [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
34. Rangan AM, O'Connor S, Giannelli V, Yap ML, Tang LM, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): Comparison of a Mobile Phone Digital Entry App for Dietary Data Collection With 24-Hour Dietary Recalls. JMIR Mhealth Uhealth 2015 Oct 27;3(4):e98 [FREE Full text] [doi: 10.2196/mhealth.4613] [Medline: 26508282]
35. Eldridge AL, Piernas C, Illner A, Gibney MJ, Gurinović MA, de Vries JHM, et al. Evaluation of New Technology-Based Tools for Dietary Intake Assessment-An ILSI Europe Dietary Intake and Exposure Task Force Evaluation. Nutrients 2018 Dec 28;11(1):e55 [FREE Full text] [doi: 10.3390/nu11010055] [Medline: 30597864]
36. Food & You. 2017. The app MyFoodRepo URL: https://www.foodandyou.ch/en/my-food-repo [accessed 2020-03-19]
37. King's College London. FODMAP App for People with Irritable Bowel Syndrome URL: https://www.kcl.ac.uk/research/work-with-kings/ip-licensing/fodmap-app [accessed 2020-03-19]
38. myfood24. Frequently Asked Questions URL: https://www.myfood24.org/faqs [accessed 2020-04-16]
39. National Cancer Institute. Dietary Assessment Primer URL: https://dietassessmentprimer.cancer.gov/ [accessed 2020-03-19] [WebCite Cache ID 78YngCDBy]
40. Ferrara G, Kim J, Lin S, Hua J, Seto E. A Focused Review of Smartphone Diet-Tracking Apps: Usability, Functionality, Coherence With Behavior Change Theory, and Comparative Validity of Nutrient Intake and Energy Estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
41. MyFitnessPal. Some food information in the database is inaccurate. Can I edit it? URL: http://www.webcitation.org/78XUneYIH [accessed 2020-03-19] [WebCite Cache ID 78XUneYIH]
42. Maringer M, Van't Veer P, Klepacz N, Verain MCD, Normann A, Ekman S, et al. User-documented food consumption data from publicly available apps: an analysis of opportunities and challenges for nutrition research. Nutr J 2018 Jun 09;17(1):59 [FREE Full text] [doi: 10.1186/s12937-018-0366-6] [Medline: 29885653]
43. MyFitnessPal. MyFitnessPal Account Security Issue: Frequently Asked Questions URL: https://content.myfitnesspal.com/security-information/FAQ.html [accessed 2020-03-19] [WebCite Cache ID 78XUy3kkf]
44. Nikolaou CK, Lean MEJ. Mobile applications for obesity and weight management: current market characteristics. Int J Obes (Lond) 2017 Jan;41(1):200-202. [doi: 10.1038/ijo.2016.186] [Medline: 27780974]
45. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile Apps for Weight Management: A Scoping Review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]

46. MyFitnessPal. 11 Things You'll Learn Using MyFitnessPal URL: https://blog.myfitnesspal.com/11-things-youll-learn-start-counting-calories-mfp/ [accessed 2020-04-16]
47. Lose It!. Incorrect Food Data URL: http://help.loseit.com/hc/en-us/articles/115007309167-Inaccurate-food-data [accessed 2020-03-19] [WebCite Cache ID 78XV1HN7X]
48. Rhyner D, Loher H, Dehais J, Anthimopoulos M, Shevchik S, Botwey RH, et al. Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study. J Med Internet Res 2016 May 11;18(5):e101 [FREE Full text] [doi: 10.2196/jmir.5567] [Medline: 27170498]
49. Pendergast FJ, Ridgers ND, Worsley A, McNaughton SA. Evaluation of a smartphone food diary application using objectively measured energy expenditure. Int J Behav Nutr Phys Act 2017 Mar 14;14(1):30 [FREE Full text] [doi: 10.1186/s12966-017-0488-9] [Medline: 28288657]
50. Cade JE, Warthon-Medina M, Albar S, Alwan NA, Ness A, Roe M, DIET@NET consortium. DIET@NET: Best Practice Guidelines for Dietary Assessment in Health Research. BMC Med 2017 Nov 15;15(1):202 [FREE Full text] [doi: 10.1186/s12916-017-0962-x] [Medline: 29137630]
51. Cade J. Measuring diet in the 21st century: use of new technologies. Proc Nutr Soc 2017 Aug;76(3):276-282. [doi: 10.1017/S0029665116002883] [Medline: 27976605]
52. Scott K, Richards D, Adhikari R. A Review and Comparative Analysis of Security Risks and Safety Measures of Mobile Health Apps. AJIS 2015 Nov 22;19:1-18. [doi: 10.3127/ajis.v19i0.1210]
53. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
54. Privacy Rights Clearinghouse. Mobile Health and Fitness Apps: What Are the Privacy Risks? URL: https://privacyrights.org/consumer-guides/mobile-health-and-fitness-apps-what-are-privacy-risks [accessed 2020-04-16] [WebCite Cache ID 78XV5ISHH]
55. Teixeira V, Voci SM, Mendes-Netto RS, da Silva DG. The relative validity of a food record using the smartphone application MyFitnessPal. Nutr Diet 2018 Apr;75(2):219-225. [doi: 10.1111/1747-0080.12401] [Medline: 29280547]
56. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. J Hum Nutr Diet 2017 Aug;30(4):439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]

Abbreviations

DAT: dietary assessment tool
FCD: food composition database

Edited by G Eysenbach; submitted 24.07.19; peer-reviewed by V Teixeira, C Boushey, D Pougheon Bertrand; comments to author 26.08.19; revised version received 13.12.19; accepted 16.12.19; published 18.06.20.

Please cite as:
Khazen W, Jeanne JF, Demaretz L, Schäfer F, Fagherazzi G
Rethinking the Use of Mobile Apps for Dietary Assessment in Medical Research
J Med Internet Res 2020;22(6):e15619
URL: https://www.jmir.org/2020/6/e15619
doi:10.2196/15619
PMID:32554383

Viewpoint

Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians

Onur Asan[1], PhD; Alparslan Emrah Bayrak[1], PhD; Avishek Choudhury[1], MSc

[1] School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

Corresponding Author:
Onur Asan, PhD
School of Systems and Enterprises
Stevens Institute of Technology
1 Castle Terrace
Hoboken, NJ, 07030
United States
Phone: 1 2012163666
Email: oasan@stevens.edu

Abstract

Artificial intelligence (AI) can transform health care practices with its increasing ability to translate the uncertainty and complexity in data into actionable—though imperfect—clinical decisions or suggestions. In the evolving relationship between humans and AI, trust is the one mechanism that shapes clinicians' use and adoption of AI. Trust is a psychological mechanism to deal with the uncertainty between what is known and unknown. Several research studies have highlighted the need for improving AI-based systems and enhancing their capabilities to help clinicians. However, assessing the magnitude and impact of human trust on AI technology demands substantial attention. Will a clinician trust an AI-based system? What are the factors that influence human trust in AI? Can trust in AI be optimized to improve decision-making processes? In this paper, we focus on clinicians as the primary users of AI systems in health care and present factors shaping trust between clinicians and AI. We highlight critical challenges related to trust that should be considered during the development of any AI system for clinical use.

(J Med Internet Res 2020;22(6):e15154) doi:10.2196/15154

KEYWORDS

human-AI collaboration; trust; technology adoption; FDA policy; bias; health care

Introduction

Artificial intelligence (AI), which has been introduced as a technology to improve decision-making involving uncertainty and complexity in systems, has the potential to transform health care practices [1,2]. The role of humans in the practical applications of AI is often overlooked. The development of automated systems to augment human decision-making dates to the 1950s with the Fitts list, which identifies the complementary capabilities of humans and automated systems [3]. The Fitts list includes 11 statements asserting that humans are better at detection, perception, judgment, induction, improvisation, and long-term memory, while automated systems are better at power/speed, computation, replication, simultaneous operation, and short-term memory [4]. Several studies have shown that automated systems may or may not improve human decision-making, depending on whether or how human factors are accounted for in their design [5].

As AI is rapidly developing, unlike other technologies, there is an absence of a clear definition of the process, functioning, and role of AI [6]. Trust is a crucial factor influencing interactions between human beings, including their interactions with AI. Understanding the trust dynamics between AI and humans is crucial, particularly in the field of healthcare, where life is at risk. In this paper, we discuss the impact of trust on the dynamic interactions between AI and clinicians, highlight the key factors that influence trust relationships and identify key challenges and future research directions in the health care domain. While the users of AI systems can be diverse, including patients and insurance providers, the focus of this paper is limited to the domain experts in healthcare, ie, clinicians. We acknowledge that trust relationships could significantly differ for patients and insurance providers.

Definitions

What Is AI?

The term AI has been used in many ways in computer science, engineering, and healthcare. Broadly, it can be defined as a computer program that can make intelligent decisions [7]. This definition includes computer programs that operate with predefined rules and data-driven models. This paper distinguishes these two by referring to the former as *automation*, which can be used to make well-defined and repetitive decisions. While automated systems have been used to augment or replace human operation in health care [8], the generalizability of the process and existence of intelligence in automation can be questionable. The focus of this research is the latter, which is a process to make health care decisions using a mathematical model built on prior or real-time data. Existing literature in the machine learning field provides relatively successful methods to train such mathematical models that learn useful knowledge from data [9-11]. In this paper, we refer to *a computer process that algorithmically makes optimal decisions based on multiple criteria using one or more machine learning-based models* as AI. While trust has been studied in the context of automation [12], the deterministic (ie, consistently providing the same output for a particular input) and relatively predictable nature of automation is an important distinction from our definition of AI, which has implications for trust research.

What Is Trust?

Interpersonal trust is a human belief (or referred to as an attitude in some sources [12]) that is broadly defined based on three main dimensions, namely, benevolence, integrity, and ability [13]. This attitude may be intrinsically formed based on the user's own experience with the system of interest or may stem from an extrinsic source such as the reputation of the system in the user's social circle [14]. Studies highlight some differences between interpersonal trust and trust in technical systems because technical systems may lack intentionality, which is relevant to honesty and benevolence [12]. A user's perception of an AI system's ability remains a significant dimension for trust in AI systems, which depends on the quality of the input data, the mathematical problem representation, and the algorithms used in the decision-making. The level of trust in AI has a significant impact on how much users rely on AI [12], and hence the efficacy of health care decisions. However, the level of trust in AI may not necessarily have a positive correlation with clinical or patient outcomes.

AI in Health Care

AI has shown significant potential in the area of mining medical records, designing treatment plans [15], robotics mediated surgeries [16], medical management and supporting hospital operations, clinical data interpretation [17], clinical trial participation [18,19], image-based diagnosis [20], preliminary diagnosis [21], virtual nursing [22], and connected health care devices [23]. In addition to these applications, significant investments in AI research [24], as well as recent efforts on regulating the use of AI in the medical domain [24], suggest that AI could become an essential technology to assist decision-making in the medical domain in the foreseeable future.

AI research in health care poses unique challenges compared to other technical domains. Physical system models mathematically describe the underlying technical behavior in engineering applications. However, the lack of such quantitative models in many health care applications such as medical diagnostics (eg, the precise relationships between diseases and their causes) creates a significant challenge. The responses from clinicians for the same clinical cases vary significantly. Therefore, it would be a challenge to train AI-based tools on the subjective responses that carry over individual biases from clinicians without any knowledge of the ground truth. Also, AI research must account for the distinct medical problem characteristics across different health care applications. It may not be possible to generalize a process to train a mathematical model for an AI tailored to the needs of cancer applications to cardiovascular applications, for instance. Further, vulnerable groups such as minorities and patients with disabilities may not be sufficiently represented in the data, and their needs may not be adequately accounted for if these groups are not carefully considered during the design of the AI system. A customized AI process might be necessary for each application depending on the type and amount of data available, the target patient population, the amount of variability and useful information in the data, and the nature of the health care decisions to be made.

Further, health care is a highly regulated space where developing an adaptive AI technology to meet regulatory requirements is an additional challenge. The US Food and Drug Administration (FDA) categorizes Software into three classes: (a) Software as a Medical Device (SaMD), (b) software in a medical device, and (c) software used in the manufacture or maintenance of a medical device. FDA defines SaMD as "… AI/ML-based Software, when intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions, are medical devices under the FD&C Act and called Software as a Medical Device" [25]. SaMD ranges from smartphone applications to view radiologic images for diagnostic purposes to Computer-Aided Detection software to post-processing of images to detect breast cancer [26]. FDA has approved several AI-based SaMDs with "locked" algorithms that generate the same result each time for the same input; these algorithms are adaptable but require a manual process for the updates [25,27].

Unlike the standard SaMD model, an adaptive algorithm changes its behavior using a definitive learning process without requiring any manual input. An adaptive algorithm might generate different outputs each time a given set of inputs is received due to learning and updating. A credible validation and verification plan that ensures safe and reliable operation under adaptive behaviors must be a part of the AI design.

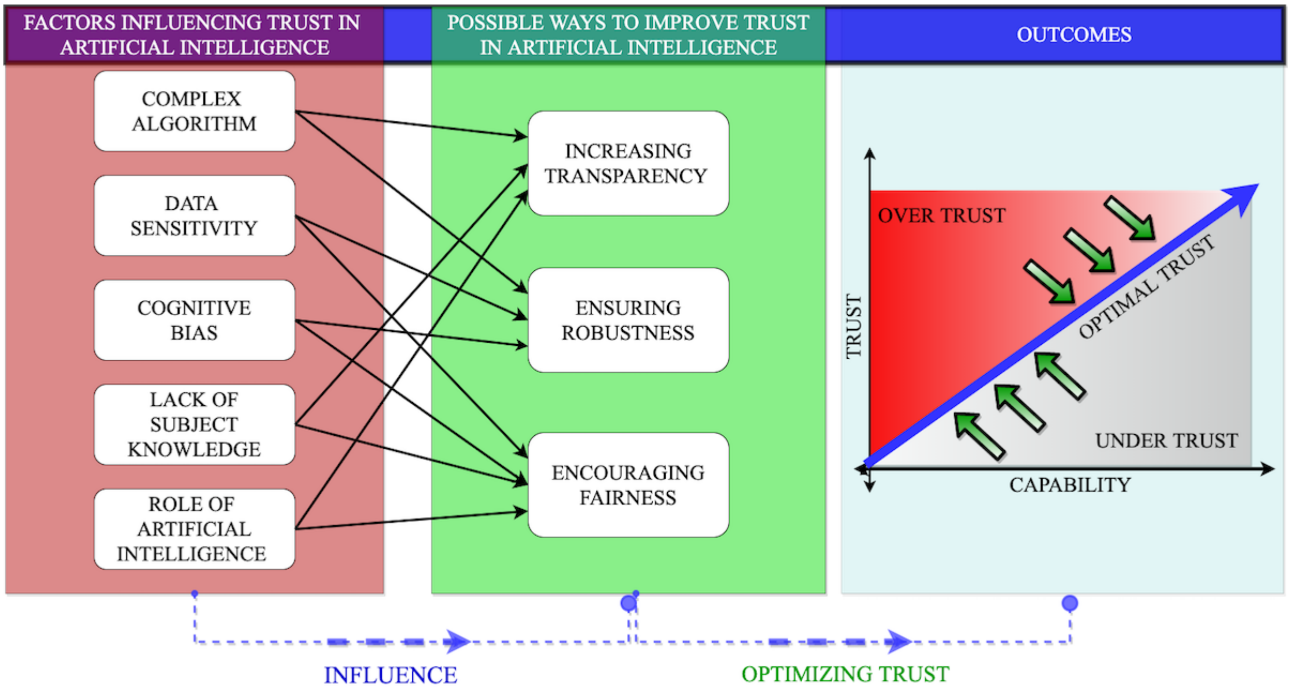
AI in health care has two potential advantages to human performance. First, AI can learn from big data (such as incommunicable silos of unstructured information stored in an electronic health record) more efficiently than clinicians. A successful AI system can efficiently extract relevant information from offline or real-time data to assist in improving organizational performance and help clinicians in making informed decisions in real time. Second, AI systems can perform predefined tasks with higher precision. AI can be in a continuous active state without compromising its performance—it does not suffer from burnout as humans do. This feature of AI technology has the potential to revolutionize complicated surgeries. The Da Vinci robotic surgical system can mimic a surgeon's hand movements with greater precision [28]. Further, existing applications of AI in various domains such as AlphaStar (an AI bot that outperforms an expert player in a video game) and LYNA (an AI capable of detecting breast cancer using images from lymph node biopsies) report successful outcomes comparable to human decision-making [29,30].

There are limitations of AI that might restrict its application in life-critical areas such as healthcare. AI methods require data inputs to be in a structured form, which limits the type of information that can be provided for medical decisions. Even the deep learning methods, which can find a suitable mathematical representation from a given dataset automatically, are trained for a given input type (eg, medical image data) and, once developed, they cannot accept another input type (eg, statements from patients). Also, AI methods generally lack "common sense," making them unable to identify simple mistakes in data or decisions that would otherwise be obvious to a human being [31]. Therefore, there is significant potential and need for improvement by combining the intuitive and analytical thinking of medical experts [32] with the computational power of AI in a proper human-AI collaboration architecture.

Trust in Human-AI Collaboration

Advances in the capabilities of AI will expand the role of this technology from the automation of repetitive and well-defined tasks to guiding decision-making under uncertainty that is currently done exclusively by medical professionals. As health care providers rely more on AI, a proper trust relationship, also referred to as *calibrated* trust [33], becomes a requirement for effective decisions. The deterministic and relatively predictable nature of a typical rule-based software system is an essential factor contributing to the development of user trust. The resulting behavior of a deterministic system can entirely be determined by its initial state and inputs. However, the nondeterministic nature of AI, where an algorithm can exhibit different behaviors in different iterations for the same input, might introduce unique dimensions to the concept of trust. Figure 1 presents our overview of some important factors influencing trust in AI for health care, possible ways to improve trust relationships, and their impact on trust. Note that the purpose of the figure is not to provide an exhaustive list but rather to highlight important issues relevant to trust in AI for health care applications.

Figure 1. Human factors and trust in artificial intelligence.



Currently, a lack of trust in the AI systems is a significant drawback in the adoption of this technology in healthcare. Trust in AI can be influenced by several human factors such as user education, past experiences, user biases, and perception towards automation, as well as properties of the AI system, including controllability, transparency, and complexity of the model, associated risks, and many others. Among these factors, reliability, which refers to whether the AI technology can perform a task predictably and consistently [34], might be particularly concerning in health care due to the changes in the reliability of AI in the presence of new data [35]. The reliability of an AI technology is conditioned on the user and input data. Considering that an AI system might be trained with insufficient and subjective data from multiple sources, AI could generate biased or overfitted outcomes of which the clinical user might not be aware. These concerns hinder the performance of this technology [36], thus deterring the user's trust and acceptance of AI systems.

It is also important to note that *maximizing the user's trust* does not necessarily yield the best decisions from a human-AI collaboration. When trust is at maximum, the user accepts or believes all the recommendations and outcomes generated by the AI system. While in some applications, AI can outperform human decision-making by incorporating data from multiple sources [29,30], the limitations above suggest that unnecessarily high trust in AI may have catastrophic consequences, especially in life-critical applications. Therefore, our study supports the concept of *optimal trust* in which both humans and AI each have some level of skepticism regarding the other's decisions since both are capable of making mistakes. The amount of skepticism necessary for the most accurate clinical decisions depends on the capability of the human user and the AI system. The development of AI must incorporate mechanisms that will establish and maintain a properly balanced, optimal level of trust from and to the user that matches the capability of the AI system [12].

Developing a healthy trust relationship in human-AI collaboration is a challenge due to the uncertain behavior of a knowledge-based evolving AI system. We posit that the following factors should be incorporated into the development of AI to achieve an optimal level of trust: fairness, transparency, and robustness (Figure 1). It is often assumed that algorithmic decision-making might lead to fairer and more robust outcomes than human judgment [37]. However, algorithms intrinsically discriminate and assign a weight to some factors over others. Moreover, the properties of machine learning algorithms bear the risk to reflect and aggravate underlying data bias, which might unfairly affect members of protected groups based on sensitive categories such as gender, race, and sexual orientation. For instance, a study [38] that implemented machine learning to detect skin cancer used less than 5% data from a dark-skinned patient population to train the model, potentially leading to bias against dark-skinned patients [39]. According to the UK International Commissioner's Office, AI fairness depends on the effects of the data processing on users (care providers and patients), their expectations as to how their data will be used (clinical and personal data) and the level of transparency provided [40,41] (sharing with other care providers and insurance). Fairness in an AI process is concurrent with bias minimization. Bias, a mismatch between the distribution of training data and a preferred fair distribution, can yield unfair outcomes (prediction/classification). In order to establish fairness in AI systems, biases originating from the subjective responses of clinicians should be identified and curtailed during validation and verification.

Achieving fairness through awareness or transparency can also improve trust [42,43]. Facilitating secured access to patient data can improve transparency [44] in the data and allow patients to validate their information. In our view, transparency fosters an understanding not only of the working principles of AI algorithms but also the algorithmic biases and biases due to underrepresentation. Depending on data availability and quality, an AI system might perform remarkably well at some tasks while performing poorly in others. Revisiting the 'skin cancer example' cited above [38], if clinicians are informed about the bias in the training data (eg, underrepresentation of minorities), then it will be easier for them to identify the suitable patient population (eg, Caucasians) on which the AI algorithm can be implemented. Explainable AI is another route to AI transparency that might help clinicians arrive at clinically meaningful explanations about the outcome of AI applications and make informed judgments. However, there are tradeoffs between the explainability and sophistication of an AI algorithm. Explainable AI models such as decision trees intuitively tend to be simple and might not be accurate, especially when dealing with big and complex data. Determining the precise balance between explainability and sophistication is crucial to enhance trust in AI. Finally, robustness is the sensitivity of the decisions made by the AI models to the input dataset. Poor robustness can lead to significant changes in the outcome of an AI model with small perturbations in the input data. Insufficient or erroneous data can impede AI robustness. Proofing AI models against such volatilities can help to build trust in AI systems.

Future Research Directions

The efficacy of human-AI collaboration is not only a function of the accuracy of the underlying mathematical process behind the AI system but also human factors, including trust. A holistic approach recognizing health care as a dynamic socio-technical system in which sub-elements interact with each other is necessary to understand trust relationships in human-AI collaboration. For instance, trust in AI systems might be affected by organizational policies, culture, specific tasks assigned to the health care providers, other similar computational tools used by the providers, providers' interaction with other individuals such as patients and other providers, as well as internal and external environmental factors. This viewpoint is consistent with and complementary to the research roadmaps proposed in the systems engineering literature on AI [45]. Applying human factors methodologies such as the SEIPS model [46] to the health care domain can assist researchers in capturing the entire socio-technical work system. These holistic human factors models provide a useful conceptual framework for researchers to capture contemporary and dynamic issues relevant to trust modeling in healthcare.

Second, recent approvals of AI algorithms reveal the limitations of existing regulatory standards. In early 2018, the FDA authorized the WAVE Clinical Platform, an early-warning system that utilizes vital sign data to identify high-risk patients [35]. FDA standards, designed for traditional rule-based algorithms, do not apply to advanced AI systems whose predictive performance might change when exposed to new data [35]. To measure the impact of AI systems, FDA should benchmark the predictive model and ensure clinically meaningful outcomes. As it does for drug approvals, FDA should rigorously confirm and test surrogate endpoints in potential evaluations of AI systems. Recently, the FDA announced that it is developing a framework for regulating AI systems that self-update on new data and seeking comments on how to regulate self-updating algorithms [47].

AI raises profound concerns regarding medical responsibility. Under current practice, clinicians are only responsible when they deviate from the standard care procedure for a given health condition (irrespective of patient health outcome) [48]. However, clinicians will be held responsible if they follow the AI recommendation when it is different from the standard care process and negatively affects patient health outcomes [48]. What would this mean for an AI system and trust between AI and users? Clinicians using AI systems are expected to use them for decision support, not as a replacement for trained clinical decision-making. In this sense, clinicians are still responsible for any medical errors that may occur as humans remain the final decision-makers. Then, in what capacity can AI assist clinicians and will clinicians be able to use and assess the reliability of an AI system? The influence of these factors on clinicians' trust in AI applications needs further study.

Considering the limitations of both human cognition and AI approaches, a quantitative measure for the optimal level of trust between clinicians and AI systems to make the most accurate and reliable clinical decisions remains unknown. Linking this optimal level of trust to specific design attributes in an AI system is another unknown. An analysis of that problem should account for the individual human factors specific to the user of the system, including the sizeable aleatory variability associated with it and continuously evolving capabilities of AI methods. The results of such an analysis should inform the regulatory policy decisions.

Finally, trust in AI is expected to have completely different characteristics for patients compared to clinicians. First, patients generally do not have expertise in the medical field, as opposed to clinicians. Further, patients, regardless of whether they are the users of the AI system or not, will be directly impacted by the clinical decisions (or suggestions) that the AI system provides. With the increased patient involvement in a patient-centered health care model, especially educated patients might question clinicians' decisions and want to be informed whether the decisions are based on AI recommendations or not. If AI systems are an essential part of this shared decision-making process between patients and clinicians, trust relationships between patients and AI systems deserve further in-depth research.

Conflicts of Interest

None declared

References

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
2. Tekkeşin A. Artificial Intelligence in Healthcare: Past, Present and Future. Anatol J Cardiol 2019 Oct;22(Suppl 2):8-9 [FREE Full text] [doi: 10.14744/AnatolJCardiol.2019.28661] [Medline: 31670713]
3. de Winter JCF, Dodou D. Why the Fitts list has persisted throughout the history of function allocation. Cogn Tech Work 2011 Aug 25;16(1):1-11. [doi: 10.1007/s10111-011-0188-1]
4. Fitts P. Human engineering for an effective air-navigationtraffic-control system. Fitts PM. editor. Oxford, England: National Research Council, Div. of; 1951.
5. Parasuraman R, Riley V. Humans and Automation: Use, Misuse, Disuse, Abuse. Hum Factors 2016 Nov 23;39(2):230-253. [doi: 10.1518/001872097778543886]
6. Siau K, Wang W. Building Trust in Artificial Intelligence, Machine Learning, and Robotics. Cutter Business Technology Journal 2018;31(2):47-53.
7. McCarthy J, Hayes P. Some philosophical problems from the standpoint of artificial intelligence. Readings in nonmonotonic reasoning: Morgan Kaufmann Publishers Inc; 1987:26.
8. Nundy S, Montgomery T, Wachter RM. Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence. JAMA 2019 Aug 13;322(6):497. [doi: 10.1001/jama.2018.20563]
9. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan 7;25(1):24-29. [doi: 10.1038/s41591-018-0316-z]
10. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2017 May 06:1236-1246. [doi: 10.1093/bib/bbx044] [Medline: 28481991]
11. Wiens J, Shenoy ES. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clin Infect Dis 2018 Jan 06;66(1):149-153 [FREE Full text] [doi: 10.1093/cid/cix731] [Medline: 29020316]
12. Lee JD, See KA. Trust in Automation: Designing for Appropriate Reliance. hfes 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50.30392]
13. Mayer RC, Davis JH, Schoorman FD. An Integrative Model of Organizational Trust. The Academy of Management Review 1995 Jul;20(3):709. [doi: 10.2307/258792]
14. Han Yu, Zhiqi Shen, Leung C, Chunyan Miao, Lesser VR. A Survey of Multi-Agent Trust Management Systems. IEEE Access 2013;1:35-50. [doi: 10.1109/access.2013.2259892]
15. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, et al. Correction: World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLoS Med 2015 Dec;12(12):e1001940 [FREE Full text] [doi: 10.1371/journal.pmed.1001940] [Medline: 26701262]

16. Davenport TH, Glaser J. Just-in-time delivery comes to knowledge management. Harv Bus Rev 2002 Jul;80(7):107-11, 126. [Medline: 12140850]
17. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
18. Fink E, Kokku P, Nikiforou S, Hall L, Goldgof D, Krischer J. Selection of patients for clinical trials: an interactive web-based system. Artificial Intelligence in Medicine 2004;31(3):241-254 [FREE Full text] [doi: 10.1016/j.artmed.2004.01.017]
19. Beck JT, Rammage M, Jackson GP, Preininger AM, Dankwa-Mullan I, Roebuck MC, et al. Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center. JCO Clin Cancer Inform 2020 Jan;4:50-59 [FREE Full text] [doi: 10.1200/CCI.19.00079] [Medline: 31977254]
20. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. Comput Struct Biotechnol J 2019;17:995-1008 [FREE Full text] [doi: 10.1016/j.csbj.2019.07.001] [Medline: 31388413]
21. Zhao M, Jiang Y. Great expectations and challenges of artificial intelligence in the screening of diabetic retinopathy. Eye 2019 Dec 11;34(3):418-419. [doi: 10.1038/s41433-019-0629-2]
22. Peete R, Majowski K, Lauer L, Jay A. Artificial Intelligence in Healthcare. Artificial Intelligence and Machine Learning for Business for Non-Engineers 2019:96. [doi: 10.1201/9780367821654-8]
23. Durkin K. Artificial Intelligence-driven Smart Healthcare Services, Wearable Medical Devices, and Body Sensor Networks. Am. J. Med. Res 2019;6(2):37. [doi: 10.22381/ajmr6220195]
24. White House American artificial intelligence initiative year one annual report. In: policy Oosat, editor. USA executive office of the president of the United States; 2020:E-29.
25. FDA. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based Software as a Medical Device (SaMD) URL: https://www.regulations.gov/document?D=FDA-2019-N-1185-0001 [accessed 2020-04-01]
26. FDA. Food and Drug Administration. 2017. What are examples of Software as a Medical Device? : U URL: https://www.fda.gov/medical-devices/software-medical-device-samd/what-are-examples-software-medical-device [accessed 2020-04-01]
27. Networks Inc. K. Statement from FDA Commissioner Scott Gottlieb, M.D. on steps toward a new, tailored review framework for artificial intelligence-based medical devices. CMR 2019 Apr 02:1. [doi: 10.31525/fda2-ucm635083.htm]
28. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019 Jul;8(7):2328-2331 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_440_19] [Medline: 31463251]
29. DeepMind. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II URL: https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii [accessed 2020-05-05]
30. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial Intelligence–Based Breast Cancer Nodal Metastasis Detection: Insights Into the Black Box for Pathologists. Archives of Pathology & Laboratory Medicine 2019 Jul;143(7):859-868. [doi: 10.5858/arpa.2018-0147-oa]
31. Goodfellow I, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. 2015 May 07 Presented at: Proceeding of 3rd International Conference on Learning Representations; May ; San Diego; 2015; San Diego p. 7-9.
32. Evans JSBT. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. Annu. Rev. Psychol 2008 Jan;59(1):255-278. [doi: 10.1146/annurev.psych.59.103006.093629]
33. Hoffman RR, Johnson M, Bradshaw JM, Underbrink A. Trust in Automation. IEEE Intell. Syst 2013 Jan;28(1):84-88. [doi: 10.1109/mis.2013.24]
34. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans. Manage. Inf. Syst 2011 Jun;2(2):1-25. [doi: 10.1145/1985347.1985353]
35. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science 2019 Feb 21;363(6429):810-812. [doi: 10.1126/science.aaw0029]
36. AlHogail A. Improving IoT Technology Adoption through Improving Consumer Trust. Technologies 2018 Jul 07;6(3):64. [doi: 10.3390/technologies6030064]
37. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. International Journal of Law and Information Technology 2019;27(2):a. [doi: 10.1093/ijlit/eaz004]
38. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
39. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. Nature 2018 Jul;559(7714):324-326. [doi: 10.1038/d41586-018-05707-8] [Medline: 30018439]
40. Office I. ICO. 2019. Guide to data protection URL: https://ico.org.uk/media/for-organisations/guide-to-data-protection-1-1.pdf [accessed 2020-04-04]
41. Butterworth M. The ICO and artificial intelligence: The role of fairness in the GDPR framework. Computer Law & Security Review 2018:257-268 [FREE Full text] [doi: 10.1016/j.clsr.2018.01.004]
42. Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness through Awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. Cambridge, Massachusetts: Association for Computing Machinery; 2012

Presented at: 3rd Innovations in Theoretical Computer Science Conference; 2012; Cambridge. [doi: 10.1145/2090236.2090255]

43. Feldman M, Friedler S, Moeller J, Scheidegger C, Venkatasubramanian S. Certifying and Removing Disparate Impact. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Sydney, NSW. Australia: Association for Computing Machinery; 2015 Presented at: 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015; Sydney. [doi: 10.1145/2783258.2783311]
44. Ferreira A, Lenzini G. Springer. Can Transparency Enhancing Tools Support Patient?s Accessing Electronic Health Records? New Contributions in Information Systems and Technologies: Springer; 2015:1121-1132.
45. McDermott T, DeLaurentis D, Beling P, Blackburn M, Bone M. AI4SE and SE4AI: A Research Roadmap. INSIGHT 2020 Mar;23(1):8-14. [doi: 10.1002/inst.12278]
46. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013 Nov;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
47. FDA. Artificial Intelligence and Machine Learning in Software as a Medical Device URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-devicefiles/7614/artificial-intelligence-and-machine-learning-software-medical-device.html [accessed 2020-05-01]
48. Price WN, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. JAMA 2019 Oct 04:1765-1766. [doi: 10.1001/jama.2019.15064] [Medline: 31584609]

Abbreviations

AI: artificial intelligence
FDA: Food and Drug Administration
SaMD: Software as a Medical Device

Edited by G Eysenbach, Q Zeng; submitted 24.06.19; peer-reviewed by E Chiou, D Walker, K Fortuna; comments to author 24.02.20; revised version received 12.05.20; accepted 03.06.20; published 19.06.20.

Please cite as:
Asan O, Bayrak AE, Choudhury A
Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians
J Med Internet Res 2020;22(6):e15154
URL: http://www.jmir.org/2020/6/e15154/
doi:10.2196/15154
PMID:32558657

Viewpoint

Artificial Intelligence and Big Data in Diabetes Care: A Position Statement of the Italian Association of Medical Diabetologists

Nicoletta Musacchio[1], MD; Annalisa Giancaterini[2], MD; Giacomo Guaita[3], MD; Alessandro Ozzello[4], MD; Maria A Pellegrini[1,5], MD; Paola Ponzani[6], MD; Giuseppina T Russo[7], MD; Rita Zilich[8], MD; Alberto de Micheli[9], MD

[1]Italian Association of Diabetologists, Rome, Italy
[2]Diabetology Service, Muggiò Polyambulatory, Azienda Socio Sanitaria Territoriale, Monza, Italy
[3]Diabetology, Endocrinology and Metabolic Diseases Service, Azienda Tutela Salute Sardegna-Azienda Socio Sanitaria Locale, Carbonia, Italy
[4]Departmental Structure of Endocrine Diseases and Diabetology, Azienda Sanitaria Locale TO3, Pinerolo, Italy
[5]New Coram Limited Liability Company, Udine, Italy
[6]Operative Unit of Diabetology, La Colletta Hospital, Azienda Sanitaria Locale 3, Genova, Italy
[7]Department of Clinical and Experimental Medicine, University of Messina, Messina, Italy
[8]Knowledge Management - Mix-x, Milano, Italy
[9]Associazione dei Cavalieri Italiani del Sovrano Militare Ordine di Malta, Genova, Italy

Corresponding Author:
Nicoletta Musacchio, MD
Italian Association of Diabetologists
Viale delle Milizie, 96
Rome, 00192
Italy
Phone: 39 02 5799 ext 4304
Email: nicoletta.musacchio@gmail.com

Abstract

Since the last decade, most of our daily activities have become digital. Digital health takes into account the ever-increasing synergy between advanced medical technologies, innovation, and digital communication. Thanks to machine learning, we are not limited anymore to a descriptive analysis of the data, as we can obtain greater value by identifying and predicting patterns resulting from inductive reasoning. Machine learning software programs that disclose the reasoning behind a prediction allow for "what-if" models by which it is possible to understand if and how, by changing certain factors, one may improve the outcomes, thereby identifying the optimal behavior. Currently, diabetes care is facing several challenges: the decreasing number of diabetologists, the increasing number of patients, the reduced time allowed for medical visits, the growing complexity of the disease both from the standpoints of clinical and patient care, the difficulty of achieving the relevant clinical targets, the growing burden of disease management for both the health care professional and the patient, and the health care accessibility and sustainability. In this context, new digital technologies and the use of artificial intelligence are certainly a great opportunity. Herein, we report the results of a careful analysis of the current literature and represent the vision of the Italian Association of Medical Diabetologists (AMD) on this controversial topic that, if well used, may be the key for a great scientific innovation. AMD believes that the use of artificial intelligence will enable the conversion of data (descriptive) into knowledge of the factors that "affect" the behavior and correlations (predictive), thereby identifying the key aspects that may establish an improvement of the expected results (prescriptive). Artificial intelligence can therefore become a tool of great technical support to help diabetologists become fully responsible of the individual patient, thereby assuring customized and precise medicine. This, in turn, will allow for comprehensive therapies to be built in accordance with the evidence criteria that should always be the ground for any therapeutic choice.

(J Med Internet Res 2020;22(6):e16922) doi:10.2196/16922

KEYWORDS

artificial intelligence; big data analytics; clinical decision making; diabetes management; health care

Introduction

The ongoing evolution in medicine and, in particular, in the field of diabetology is strongly intertwined to a series of changes and innovations [1-3]. The term “digital health” is a “container,” grouping together informatics and telecommunications that have the common objectives of diagnosis, treatment, or monitoring of diseases, maintenance of health and well-being, and support for healthy lifestyles. The US Food and Drug Administration has compiled a list of software programs with medical device functions (apps for smartphones and personal computers with diagnostic, monitoring, or therapeutic objectives), advanced business intelligence data analysis tools, artificial intelligence (AI), cloud, cyber security, and innovative health technologies [4]. These tools belong to an evolving reality; they are still unclear but they are potentially leading to challenging and promising scenarios [5,6].

The scope of this position statement is to analyze the most relevant aspects and describe the changes that have already occurred or those that will take place shortly by exploring the possibilities of application and development in the field of diabetology. Indeed, the digital world is constantly expanding and it has already become an integral part of our personal and professional life [7,8]. Smartphones, personal computers, and network access are now essential tools for almost all populations and for many diabetologists, both as individuals and professionals.

Digitalization

Almost every daily activity of the diabetologist has to deal with digitalization, from electronic medical records to imaging diagnostics, from laboratory references to the various software programs for administrative practices and certifications, and from glucose level data downloaded from glucometers to sensors for continuous glucose monitoring and insulin pumps. Through the analysis of these data, the diabetologists make therapeutic decisions, and the more complete and numerous the data, the more they need informatics tools to guide them in the analysis, help them identify specific patterns for detecting glycemic abnormalities, understand the possible causes, and adopt the appropriate therapeutic strategies to correct these abnormalities [9,10].

Data Management and Connectivity

In the early days, data collection and management for each individual patient occurred manually at the hospital by physically connecting cables or wireless devices to the local software/network. Currently, thanks to the technological evolution, these data can be transmitted automatically from the glucometers via “cloud” or on data integration platforms that collect elements from different devices and provide standardized reports [11,12]. If not properly managed, this large amount of data is likely to overwhelm both the patient and the health care professional, and the professionals’ experience in reading these reports is certainly the added value that increases their potential. Advanced informatics tools can simplify the analysis and provide suggestions to guide the clinical decisions of the physician [13]. With experience and technology coming together, the analysis time can be increasingly reduced and more appropriate and correct data-driven therapeutic decisions can be made, thanks to the greater customization strategies of the therapy [14-21].

AI and Individualized Care

In this context, the use of AI, and in particular, the use of machine learning allows an important step forward compared with the traditional data analysis techniques (ie, graphs representing a picture of reality, which are very useful and precise but are static and outdated). Through the automatic identification of specific patterns within the data and through inductive reasoning typical of the human mind, machine learning can highlight correlations that lead to “predictions,” without being programmed in advance to carry out this activity [19].

In a not too distant future, AI, thanks to algorithms that can enable learning and improving the machine’s own abilities independently, will offer effective solutions to satisfy the most disparate needs and will be able to deal with problems that today may seem insurmountable obstacles, for the benefit of the community [22]. In the field of clinical diabetology, these tools could have multiple potential implications, including the identification of new risk factors for the onset of diabetes (through the evaluation of large databases related to the general population) or revealing unsuspected subjects who are at a high risk of complications (by cross-checking clinical and administrative records of patients with diabetes) and identifying the behavioral and therapeutic variables that are most closely related to the progression of a specific complication.

There are many examples of collaboration between pharmaceutical companies, information technology companies, scientific institutes, and universities that exploit large complex datasets (the so-called big data) with the aim of improving the treatment of type 2 diabetes mellitus (T2DM) and of unraveling physiopathological mechanisms through the integration of data from biological, demographic, clinical, environmental, and genomic sources [23].

Risk Stratification and Personalized Medicine

The ability to elaborate a large and heterogeneous amount of data, even in real time, through increasingly powerful algorithms, allows us to extract knowledge and exponentially make predictive assessments on the behavior of individuals as well as, more generally, assume decisions for the entire community [24,25]. This means that the use of AI will make it possible to transform the immediate data (descriptive) into knowledge of the factors that “condition” behavior and correlations (predictive) [25-27] up to the identification of the key factors that can facilitate an improvement in the expected results (prescriptive) [28,29]. In this scenario, one of the most interesting applications would be to identify the variables that could be related to greater responsiveness to a specific drug because this approach would open the door to a truly personalized medicine that uses the right drug for the right person, with obviously greater efficacy, improved outcomes, and containment of costs. Furthermore, predictive analysis techniques based on AI could be used to identify which group of patients would require more attention and which strategies would be more effective, depending on the individual patient, thereby allowing more efficient methods of personal care with lower costs and better outcomes. This concept of risk stratification, which considers all the individual characteristics of the patient (ie, clinical aspects, genetic data, lifestyle, environmental factors), is the basis of the modern model of clinical governance called “population health management.” Today, the risk stratification of the population results from the extraction of historical expenditure data (eg, inpatient and outpatient admissions, diagnosis, pharmaceutical expenditure) and this presents the following series of biases: weak clinical validity; temporal misalignment between the extraction, the analysis, and the health status; and absence of information on the real socioeconomic condition and behavioral health. In this way, the data represent “satisfied demand” rather than “real needs.” The new challenge is the use of big data and business analytics systems to activate risk stratification models based on real health status, integrated use of multiple sources, and collaboration between professionals (ie, care team and data manager in a data mining process) [30,31].

Biotechnologies and Omics

Another aspect that characterizes the ongoing medical revolution is the development and increasing importance of biotechnologies used in genetics and in the so-called “omics” sciences. These new techniques have been showing that the complex pathophysiological processes underlying type 1 diabetes mellitus (T1DM), T2DM, and gestational diabetes mellitus are caused by disturbances in the gene expression that lead to alterations in the processes within the organs involved in glucose homeostasis [32]. The complexity of the system is exacerbated by the fact that the relative contribution of each component is highly individual. Understanding the molecular mechanisms underlying these interactions is crucial for developing new personalized prevention and treatment strategies [32,33].

Social Networks and Apps

Finally, the ever-increasing use of social networks and apps has major implications in the diabetes care. On one hand, health professionals must change the way they communicate and keep up with the times by always acquiring new technological and communication skills to deal with and manage the changes taking place, while on the other hand, the new technologies could be used as motivational and coaching tools in support of the traditional educational activities and as an alternative data source. Notably, the areas that influence the state of health are only 10% due to the medical care and health status, whereas the major role is instead played by lifestyle and behaviors as well as by genetics [34-36]. For this reason, to obtain a more precise picture of the state of health of the general population or of specific subgroups, we cannot limit ourselves to collecting and analyzing only health data, but we must consider what comes from the world of social networking and smartphone apps, despite all the known limitations and criticalities.

Evolution of the Role of Health Care Professionals

The exponential increase in knowledge and technologies, the increased complexity of the tasks, and the increasingly diversified needs of the patients are flooding health care professionals with increasingly greater tasks and aspects to be addressed and managed. It is necessary to reflect on what the true essence of medical care is and to reassess the meaning of this profession. The ability to change, juggle the new technology tools, and exploit the potential of new information technology and business analytics techniques will enable health care professionals to have a great support in their choices and reduce the time spent collecting data or using the machines so that they can concentrate more on the decision-making process, thereby rendering each intervention more effective and efficient. As a proof of concept, it has been recently reported that the costs for care of patients were steadily diminished by the active use of technology, which supplied a real time feedback about their blood glucose levels and connected them to clinical support [37]. These reductions were primarily attributable to reduced diabetes-related and office-based services, independent of the specific condition, and they were proportional to the frequency of use of the digital tools [37].

Data from Diabetology Literature

Diabetology is facing different challenges: the ever-decreasing number of diabetologists, the increasing number of patients, the reduction of visitation time, the increasing complexity of the pathology both from the clinical and welfare point of view, the difficulty in achieving the objectives, the growing burden of pathology management for health care professionals and patients, and the decreased accessibility to care and sustainability. New digital technologies and the use of AI are certainly a great opportunity. The current panorama of international scientific literature with respect to the use of big data and AI in diabetology offers various hints and in-depth analyses, which are applicable to different fields.

The Epidemiological Area

Cases of diabetes have been identified within large heterogeneous databases, and new risk factors for diabetes have been identified. An interesting multi-database retrospective study on the identification of cases of T2DM [17] strategically used the European Medical Information Framework Project database, a European project for the efficient reuse of health data for epidemiological research [38]. This database collects health information of about 52 million European citizens by using heterogeneous sources and by acting as a support for the execution of high-quality multinational observational studies, based on populations with large sample sizes and otherwise inconceivable follow-ups. Subjects with T2DM were identified using a complex algorithm strategy in 8 different health data sources, and the strengths and limitations of each data source were revealed during the creation of a model that ensures the interoperability of systems of heterogeneous electronic medical records. These efforts represent a methodological advancement for carrying out studies of multinational and multi-data sources, thereby providing sufficient information for the contextualization and correct interpretation of the results and generating transparent and reusable documentation.

Additionally, a study conducted by researchers from the University of New York and Philadelphia described a new "data-driven" approach of population health management based on the use of machine learning techniques to develop predictive models and identify risk factors for the onset of T2DM [39]. They based this model on administrative data concerning health services, pharmaceutical records, insurance databases, citizens' access to the different care facilities, and laboratory results from 4.1 million individuals over a period of 4 years and a total of 42,000 variables. This model has been able to identify new risk factors for the appearance of diabetes with a predictive probability of at least 50% higher than the models based on the known traditional risk factors, thereby avoiding the costs of performing a screening. The identification through new predictive models based on machine learning of the part of the population that is at the highest risk of developing diabetes will be able to generate clinical hypotheses for the identification of new risk factors and to implement more targeted cost-effective interventions.

Phenotyping and Risk Stratification

A study of 65,000 newly diagnosed patients with T2DM [40] estimated that 10% of the subjects absorb 68% of the resources. The task of predictive analytical methods based on big data is precisely that of identifying the 10% of the subjects most at risk, on whom the treatment needs to be intensified in order to improve the health outcomes at lower costs.

An interesting Japanese study [41] has developed a practical framework for phenotyping T2DM by using both specialized knowledge and an approach based on machine learning (in particular, the support vector machine) to develop 2 phenotyping algorithms with data extracted from electronic medical records: one with high sensitivity for screening and the other to identify the subjects for research. Both algorithms showed superior performance compared to the basic algorithms, thus suggesting the possibility of using the proposed framework to conduct appropriate research depending on the objective.

The Diagnostic Area

The use of predictive models based on AI has demonstrated the feasibility of identifying individuals with the highest probability of having undiagnosed diabetes through clinical data that can be easily obtained from different databases [42], thereby exploiting the potential of machine learning algorithms, including neural networks, as tools for diagnosing diabetes [43]. Machine learning and pattern recognition are tools of enormous interest as they are promising in improving the sensitivity and specificity of disease detection and diagnosis, and they appear to be able to reduce the potential for human error in decision-making. An exemplar application is the creation of surveillance algorithms that are able to detect diabetes and, in particular, to distinguish T1DM from T2DM by using structured electronic medical records [14]. The extraction of data from the electronic database of a 4-year long, multi-sectorial, and multi-specialist outpatient practice allowed the inclusion of approximately 700,000 patients. Possible cases of diabetes were reported using laboratory test results, diagnostic codes, and prescriptions. More cases of diabetes were captured by taking advantage of the entire range of data acquired from the records compared to that captured by the analysis of only administrative databases, thus increasing the sensitivity of the method. The application of these algorithms to electronic folders has the potential to provide timely and clinically detailed information on large numbers at low cost and nearly in real time. Electronic records will probably become increasingly important sources for the surveillance of public health and for the definition of more targeted health policies [15,44].

Field of Automatic Reporting

Diabetic retinopathy, in particular, is a chronic complication of diabetes. The automated classification of diabetic retinopathy has potential benefits such as increasing the efficiency, reproducibility, and coverage of screening programs; reducing obstacles to accessing and improving results; and providing early diagnosis and treatment. To maximize the clinical utility of the automated classification, an algorithm designed to detect specific lesions or to predict the presence of any level of diabetic retinopathy was recently developed [45]. In this study, deep learning, together with visual and pattern recognition techniques, has allowed the identification of the desired features with the highest predictive value directly from the images on a large set of labeled example data. These results show that deep neural networks can be "trained" by using large data sets and without having to specify lesion-based features to identify diabetic retinopathy or diabetic macular edema in fundus images with high sensitivity and high specificity.

Economic Field

Cost-effectiveness studies of health interventions have been performed previously. An Italian group performed a retrospective analysis through the cross-examination of large clinical and administrative databases with the aim of quantifying the relationship between health care costs attributable to diabetes and the level of glucose control [46]. The results indicated that glycemic control (expressed by hemoglobin A_{1c} [HbA_{1c}] levels) is a useful surrogate not only to estimate the odds of developing diabetes-related complications but also to estimate the costs associated with health care. The integration of administrative and clinical databases seems to be suitable to show that an appropriate management of diabetes can allow a better allocation of resources.

Expected Advantages in Diabetology

The ongoing medical revolution is strongly linked to the spread of digital health, the new software for AI, the use of big data to make more appropriate data-driven decisions, and the even greater focus on predictive, preventive, personalized, and participatory medicine [6]. Each of these elements has important repercussions in the management of complex and widespread chronic diseases such as diabetes. Table 1 summarizes the advantages and weaknesses attributed to each area of application, especially considering the challenges that diabetology faces, according to our professional experience and opinion.

Table 1. Advantages and weaknesses of the use of new technologies in diabetology (our opinion).

New technology in diabetology	Advantages	Issues
Digital data management (glucometers and continuous glucose monitoring connected to the cloud and data integration platforms)	• Support for doctors' decisions • Reduced analysis time • Graphs and images easy to understand and interpret • Correct management, supported by data, even remotely • Sharing with caregivers or family members possible • Simultaneous analysis of data from different devices • Integration of glycemic values with alternative data for better understanding (eg, carbohydrate intake, physical activity) • Possibility of intervening in the intervals between visits • Overcomes geographical barriers • Motivational tool	• Difficult to integrate with computerized clinical records • Different software programs for different devices • Time spent learning the software and gaining experience • Risk of data "flooding" the professional and the patient • Lack of significant evidences on the improvement of the outcomes • Limited number of patients currently accessing this technology • Lack of recognition for time spent and medical services • Requirement of organizational changes
Mobile app (medical device with CE[a] marking)	• Therapeutic instrument (eg, bolus calculator) • Easy visualization of data and management of corrective actions • Overview of trends over time • Greater patient involvement • Convenient for the patient • Motivational and educational support tool	• New skills and time for patient training • Reliability of the instruments
Telemedicine	• Overcomes geographical barriers • Greater accessibility to care • Reduced administrative burden (if structured) • Lower costs and inconvenience to the patient • Integration with traditional management in the clinic • Strong potential for cost reduction	• Nonrecognition of medical services • Structural difficulties • Need for institutional and organizational changes
Machine learning	• Performance of descriptive, predictive, and prescriptive analyses • Analysis of large databases of different sources that cannot be analyzed with traditional statistics • Better epidemiological risk assessment of the disease • Identification of new variables and new risk factors for the development of diabetes and its complications • Possibility of identifying the most effective patient-tailored therapeutic strategy • Minimizing adverse drug events by increasing safety • Possibility of phenotype/genotype integration	• Data quality • Heterogeneity of unstructured data • Correct use of data • Integration of data from different sources • Respect for privacy • Ethical problems • Possibility of reducing the professional skills of doctors • Replacing the professional with the machine • Difficulty in knowing and interpreting new analysis models different from the traditional clinical epidemiology of evidence-based medicine

[a] Conformité Europèene.

Importance of the Chronic Care Model

A very interesting field of application concerns the Chronic Care Model (CCM), a medical assistance model for patients with chronic diseases. The CCM is a model for sustainable chronicity [47], in which the scope is to achieve a "personalized and effective care," with an active participation of the person, integrating the different professionals involved in assistance, to concretely improve the life of the person with a chronic pathology. The CCM in the declination of its various dimensions (health care organization, delivery system design, decision support, clinical information systems, self-management support, and community resources) brings a huge amount of the so-called "real world data." Such real world data are data collected in the normal clinical practice (not from controlled clinical trials), which allow the description of the patient's care pathways through the integration of different sources, consistent with what usually happens in reality. Each step of the care and assistance pathway—from diagnosis to treatment and follow-up—generates a large amount of data and images (big data) that often reside within health facilities in separate and independent databases. To obtain an integrated view of the diagnostic-therapeutic pathways for individual patients and to trace their complexity, it is necessary that these flows are integrated.

An interesting Australian study on CCM [48] shows how the use of electronic medical records and informatics-integrated assistance can improve the management of chronic diseases. The benefits for health care professionals and service users through an accurate and timely exchange of information are better work efficiency, prevention of repetition of data and information collection, as well as a better decision-making process.

Perspective of AMD

In Italy, in total, about 5 million people have diabetes (about 1 in 12 residents), and the number will probably rise to 7 million in 15-20 years. Italian epidemiological data suggest that around 250,000 new diagnoses of T2DM and around 25,000 new diagnoses of T1DM are reported each year [49]. Since the 1970s, an articulated network of diabetes centers has developed—some operating independently as simple, complex, departmental, or territorial structures, while others with outpatient activities are attached to the Internal Medicine and Geriatrics Operative Units. The social and health relevance of diabetes has been sanctioned by a law (n. 115 of 1987) that has enhanced the role of diabetes centers and inspired numerous national and regional documents for the following 30 years and the approval of the National Diabetic Disease Plan in 2013 by the Ministry of Health. This model encompasses family doctors and a widespread network of specialist centers throughout the national territory, based on multi-professional skills (eg, diabetologist, nurse, dietitian, sometimes psychologist or podiatrist, and eventually cardiologist, nephrologist, neurologist, ophthalmologist) and regularly serving about 50% of the people with diabetes. As a result of this network, Italy has the lowest average level of HbA_{1c} and the lowest rates of chronic complications and excess mortality among the western countries [50,51]. In this regard, the role of diabetes care in reducing mortality in people with diabetes must be underlined; those who are assisted in diabetes centers have lower total and cardiovascular mortality than those who do not attend the diabetes centers [50,51].

The mission of the Italian Association of Medical Diabetologists (AMD) is to promote the professional development of its members and to ensure the continuous improvement of the quality of assistance to all people with diabetes. In ancient times, AMD had already grasped the importance of standardizing the recording of daily clinical work in an electronic folder, followed by the collection and processing of real life information on care, examinations, complications, and therapy for a qualitative interpretation of real assistance in the national territory. This intuition of the use of variables with an intrinsic clinical and professional integrated meaning, which imply the doctor's reasoning in predictive and prescriptive choices has in fact realized the first model of decisional learning, thanks to the use of statistical algorithms.

Experience of the Annals

The AMD Annals represent a periodic publication that has allowed, annually since 2006, to assess the care profiles of people with T1DM and T2DM who have been admitted in Italian diabetes care centers. A wide network of diabetology services equipped with a computerized clinical record system, used for the normal management of patients in charge, has a software program provided by AMD that allows the extraction of a standardized set of clinical information (ie, Data File AMD). The database obtained is used to calculate the quality of the care indicators, both centrally and locally. In this way, a benchmarking activity is promoted based on the comparison of one's own performance with that recorded at the national level (best performers approach). This activity reflects the professional performance of the diabetologists—it is perfectly in line with the recommendations of the National Diabetes Plan. It is very cost-effective and has produced a systematic improvement over the years of all the indicators considered [52].

In addition, the AMD Annals database represents a valuable source of observational research data. In fact, it has allowed the investigation of many key aspects over time, such as the care of the elderly patients, gender medicine, cardiovascular, renal, and hepatic issues, the appropriateness of the use of drugs, and glycemic self-control, thereby providing a realistic picture of the work of the diabetology services. Notably, the analysis of the Annals database has allowed the identification of critical areas and, therefore, the timely activation of processes of improvement, in a logic of continuous quality enhancement, that is, a process of periodical performance assessment of diabetes centers on data collection and quality of care according to standardized indicators [53-57]. It is easy, at this point, to visualize that the Annals project "photographs" the history of the evolution of clinical diabetology and has allowed, for its own conception and structuring, each individual center to self-evaluate and improve in real time. Precisely, this careful measurement of behaviors and results is capable of triggering a dynamic evolution of an entire class of professionals, thus representing an irreplaceable benefit.

With innovative and far-sighted thinking, AMD has created this database, which today represents an unparalleled heritage in a cultural civilization that is increasingly aware of the importance of information and the value of a large and accurate collection of data. In this new era, the world is being organized to take advantage of the ever-increasing databases to rely on technology and use them as assets to be interpreted to facilitate and accelerate important decisions in every field. It is clear that it is increasingly important to have a quality data collection that is increasingly "clean," and it is fundamental to highlight the need for all professionals to be trained in the culture of big data and its correct recording. In fact, given that decisions can be made on the results of data processing, it is essential to raise the problem of the truthfulness of the data on which the analyses are based, to have a measure of the reliability of the results, and to be aware of them. Further, in this area, AMD has invested time and resources in activating a process—born more than 10 years ago—of culture on the quality of the data, which involved many partners who participated in the Annals project [53-57]. This is why AMD feels ready to investigate the topic, as it has an advantageous background, and it wants to project itself in a competent and proactive way in the world of big data and AIs that represent a new cultural challenge in the scientific field.

Diabetes Intelligence, the Value of the Diabetologist: Structured Skills and Prioritized Activities

The DIA&INT (Diabetes Intelligence) is an AMD project that rises in the setting of exploring new methods of analysis of unconventional data, and it can be considered as the first experiment in analyzing unstructured data by using the business intelligence method (a precursor of machine learning systems).

In a comprehensive vision of clinical activity, AMD focuses on the value of the skills that identify a diabetologist and make it a decisive tool in the care process. To achieve this objective, AMD implemented a business intelligence project called DIA&INT [47]. This project was mainly aimed at encouraging the implementation of the CCM in an "evidence-based" way by highlighting the direct link between the activities carried out by the diabetologist and their expected outcomes in order to highlight those activities that are essential to obtain the best outcomes in modern diabetology. Furthermore, this method may contribute to extrapolating the actions that could optimize the scarce resources and it may represent a valid support for the institutional choices in the current health system revision.

Admittedly, DIA&INT has been designed to respond to the need for establishing the qualitative dimension of the diabetologist's performance and the specific contribution of the many factors participating in the clinical decision in the real world. To do this, AMD has chosen to use the following accredited tools.

1. Organizational analysis to measure and enhance the role of diabetes care with specific tools (social return on investment).
2. Data management with advanced tools (business intelligence).

The resulting "program" has been structured to standardize the information and define activities and competences, as implicitly described in the official guidelines, which are measurable and comparable, with processing methods different from those of classical epidemiology, but necessary to perform predictive and prescriptive assessments. The methodology used [47] exploited the collective intelligence of the diabetologists who participated in a survey, expressing their opinion on a complex node such as the definition of the integrated benefit of certain activities in a personalized and weighted way. The intention was to "display" how diabetologists think when carrying out clinical judgment.

Through this approach, 2 main concepts were outlined: the requirements (priority, specificity, frequency, and multidimensional analysis) and the decision-making elements for "if…then" choices based on a priori knowledge (medical, regulatory, ethical, psychosocial, etc) and the ability to propose a dynamic personalized treatment project to the person with diabetes that does not yet involve healing and that requires an active personal effort. By doing so, DIA&INT has selected the necessary skills and activities in medical practice, which have an impact on the evolution of the quality of the "health product" and the outcomes. DIA&INT produced the Core Competence curriculum of the diabetologist [58] and measured the impact of the activities on the experiences [47].

Discussion

The changes offered by technological innovation have generated an unprecedented level of data collection and processing that is destined to undergo further expansion with the new applications of robotics and augmented reality, crossing a new frontier and entering the era of big data and cognitive systems. A new category of technologies is born, which uses natural language processing and machine learning and is able to amplify and accelerate the digital transformation process to allow people and machines to interact in a more natural way, extending and enhancing cognitive skills and abilities. The possibility of extracting information that has a meaning and is functional, in fact, requires the development of sophisticated technologies and interdisciplinary skills to operate closely together. In particular, in medicine, health care systems require consistent, appropriate, and sustainable choices.

Today, the complexity of medicine certainly goes beyond the capacity of the human mind. The patients themselves are increasingly complex and we know that the long-term effectiveness of a treatment depends on variables that are no longer just "numerical," but also on other information that is difficult to structure [2]. In this framework, advances in computing power play a central role in the acquisition of knowledge. It is essential to collect and use the key information in a coherent way, out of its abundance by using effective and reliable analysis tools that are represented by the new techniques of AI already available today. These techniques recognize and use machine learning systems that are able to "extricate themselves" and learn from these immense amounts of data, even with intrinsic systems of recognition and error management. In essence, AI is a machine capable of solving problems and reproducing activities typical of human intelligence [59].

However, at this stage, it is important to remind our readers that even though the data are processed through algorithms, the final decision rests with the physician, and the fact that still a huge part of this category is reluctant to computerization should not be overlooked. In addition, more research is needed to assess how clinicians currently incorporate digital tools into their practice and the prevalence of digital platform use in various health care systems. AMD has already set the basis of culture and tools on the importance of collecting data, but now, it is necessary to go further. The strong potential of the large amount of data collected in more than 10 years of observation in the AMD Annals initiative may contain a significant "hidden" knowledge that, analyzed through different tools, may show new patterns that can help us to make better decisions in the prevention and management of diabetes and its complications. For example, this type of analysis may not only amplify our notions on the risk factors for diabetes complications but may also unravel new and unsuspected connections between them, indicating the probability with which they can affect the evolution of the disease on specific groups of subjects. Properly "trained" machine learning algorithms are able to efficiently evaluate millions of data while seeking probabilistic risk correlations, and they are not limited to tracking the spread of an epidemic but has the scope of identifying new personalized therapies. We could also imagine going further, in the increasingly realistic hypothesis of linkage between the various health care databases (administrative data, assistance process, intermediate and final results, costs, etc), which will allow us to make comprehensive assessments of the whole process of care in an even more individualized way in terms of clinical effectiveness, organizational effectiveness, sustainability, and equity.

The field of big data is already a reality and this systematic revision and position statement was made to offer the basis of knowledge for a constructive debate on the next steps that the AMD as an association and, more in general, diabetology would follow in the very next future. It is time to choose to be proactive players in this complex system and take responsibility for these new processes aimed at improving the care of people with diabetes, which is and remains our “mission.”

In conclusion, the incessant progress of these changes brings into question many established paradigms, also in the scientific sphere, and this position statement represents the first document of careful analysis of this new world, which we must make our own in a logic of constructive comparison. We must reflect on the current scenarios, ask ourselves about the effects produced by these transformations to understand the consequences in our lives induced by the automated decisions, and become able of integrating the traditional route into the new systems.

The diabetologist, more than any other health care professional, already has the right mentality of being ready to take up this innovation. The culture of the data is in our DNA, the need to phenotype the patient and personalize the care and therapeutic approach have long been our priorities, the skills to manage a complex disease such as diabetes have been refined over time—ranging from technological to communicative and from educational and andragogical to managerial abilities. Thanks to all this progress made over the years and strongly desired by AMD, we are ready for a new challenge in the management of diabetology that must see us as protagonists. The “digital diabetes” is coming, and AI and big data are opening a window on new scenarios. Today’s diabetologists must acquire new skills to be able to lead these changes, be proactive in exploiting their potential and advantages, and limit their risks and guard the essential elements of the professions of doctors and health care professionals managing patients with complex chronic conditions.

Acknowledgments

No funding was received for the preparation of this manuscript. This document was approved and validated by the following members of the National Executive Board of AMD 2015-2017: N Musacchio, D Mannino, A Botta, M Calabrese, F Calcaterra, R Candido, A Gigante, G Guaita, G La Penna, S Leotta, A Lo Presti, MC Ponziani, and M Ragonese.

Authors' Contributions

NM conceptualized and drafted the manuscript. AdM directed the “BIG DATA and AI” Position Statement group of AMD, coordinated the writing of the manuscript, and reviewed it. AG, GG, AO, MP, PP, GR, and RZ contributed to the writing, editing, and reviewing of the manuscript.

Conflicts of Interest

None declared.

References

1. Moskowitz A, McSparron J, Stone DJ, Celi LA. Preparing a New Generation of Clinicians for the Era of Big Data. Harv Med Stud Rev 2015 Jan;2(1):24-27 [FREE Full text] [Medline: 25688383]
2. Obermeyer Z, Lee TH. Lost in Thought - The Limits of the Human Mind and the Future of Medicine. N Engl J Med 2017 Sep 28;377(13):1209-1211 [FREE Full text] [doi: 10.1056/NEJMp1705348] [Medline: 28953443]
3. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA 2018 Apr 03;319(13):1317-1318. [doi: 10.1001/jama.2017.18391] [Medline: 29532063]
4. U.S. Food and Drug Administration. 2018. Digital Health Criteria URL: https://www.fda.gov/medical-devices/digital-health/digital-health-criteria [accessed 2019-10-21]
5. Iyengar V, Wolf A, Brown A, Close K. Challenges in Diabetes Care: Can Digital Health Help Address Them? Clin Diabetes 2016 Jul;34(3):133-141 [FREE Full text] [doi: 10.2337/diaclin.34.3.133] [Medline: 27621530]
6. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. JAMA 2016 Dec 13;316(22):2368-2369. [doi: 10.1001/jama.2016.17217] [Medline: 27898974]
7. Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine. Ann Intern Med 2016 Apr 19;164(8):562-563. [doi: 10.7326/M15-2970] [Medline: 26809201]
8. Scott IA. Machine Learning and Evidence-Based Medicine. Ann Intern Med 2018 Jul 03;169(1):44-46. [doi: 10.7326/M18-0115] [Medline: 29710098]

9. Eghbali-Zarch M, Tavakkoli-Moghaddam R, Esfahanian F, Sepehri MM, Azaron A. Pharmacological therapy selection of type 2 diabetes based on the SWARA and modified MULTIMOORA methods under a fuzzy environment. Artif Intell Med 2018 May;87:20-33. [doi: 10.1016/j.artmed.2018.03.003] [Medline: 29606521]
10. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
11. Nielsen KB, Lautrup ML, Andersen JKH, Savarimuthu TR, Grauslund J. Deep Learning-Based Algorithms in Screening of Diabetic Retinopathy: A Systematic Review of Diagnostic Performance. Ophthalmol Retina 2019 Apr;3(4):294-304. [doi: 10.1016/j.oret.2018.10.014] [Medline: 31014679]
12. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. Diabetes Care 2019 Aug;42(8):1593-1603. [doi: 10.2337/dci19-0028] [Medline: 31177185]
13. Hansen MM, Miron-Shatz T, Lau AYS, Paton C. Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives. Contribution of the IMIA Social Media Working Group. Yearb Med Inform 2014 Aug 15;9:21-26 [FREE Full text] [doi: 10.15265/IY-2014-0004] [Medline: 25123717]
14. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. Diabetes Care 2013 Apr;36(4):914-921 [FREE Full text] [doi: 10.2337/dc12-0964] [Medline: 23193215]
15. Makam AN, Nguyen OK, Moore B, Ma Y, Amarasingham R. Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. BMC Med Inform Decis Mak 2013 Aug 01;13:81 [FREE Full text] [doi: 10.1186/1472-6947-13-81] [Medline: 23915139]
16. Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine Learning Methods to Predict Diabetes Complications. J Diabetes Sci Technol 2018 Mar;12(2):295-302 [FREE Full text] [doi: 10.1177/1932296817706375] [Medline: 28494618]
17. Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, Egger P, et al. Identifying Cases of Type 2 Diabetes in Heterogeneous Data Sources: Strategy from the EMIF Project. PLoS One 2016;11(8):e0160648 [FREE Full text] [doi: 10.1371/journal.pone.0160648] [Medline: 27580049]
18. Raman R, Srinivasan S, Virmani S, Sivaprasad S, Rao C, Rajalakshmi R. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. Eye (Lond) 2019 Jan;33(1):97-109 [FREE Full text] [doi: 10.1038/s41433-018-0269-y] [Medline: 30401899]
19. Rodríguez-Rodríguez I, Chatzigiannakis I, Rodríguez JV, Maranghi M, Gentili M, Zamora-Izquierdo M. Utility of Big Data in Predicting Short-Term Blood Glucose Levels in Type 1 Diabetes Mellitus Through Machine Learning Techniques. Sensors (Basel) 2019 Oct 16;19(20):4482 [FREE Full text] [doi: 10.3390/s19204482] [Medline: 31623111]
20. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
21. Williams BM, Borroni D, Liu R, Zhao Y, Zhang J, Lim J, et al. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. Diabetologia 2020 Feb;63(2):419-430 [FREE Full text] [doi: 10.1007/s00125-019-05023-4] [Medline: 31720728]
22. Ridi A. Associazione Medici Diabetologi - AMD. 2017. La gestione dei Big Data: strumenti di governance e appropriatezza URL: http://aemmedi.it/wp-content/uploads/2017/07/08_Europa_ridi.pdf [accessed 2019-10-21]
23. Lo-Ciganic W, Donohue JM, Thorpe JM, Perera S, Thorpe CT, Marcum ZA, et al. Using machine learning to examine medication adherence thresholds and risk of hospitalization. Med Care 2015 Aug;53(8):720-728 [FREE Full text] [doi: 10.1097/MLR.0000000000000394] [Medline: 26147866]
24. Gong F, Chen Y, Wang H, Lu H. On building a diabetes centric knowledge base via mining the web. BMC Med Inform Decis Mak 2019 Apr 09;19(Suppl 2):49 [FREE Full text] [doi: 10.1186/s12911-019-0771-6] [Medline: 30961582]
25. Shah ND, Steyerberg EW, Kent DM. Big Data and Predictive Analytics: Recalibrating Expectations. JAMA 2018 Jul 03;320(1):27-28. [doi: 10.1001/jama.2018.5602] [Medline: 29813156]
26. Alanazi HO, Abdullah AH, Qureshi KN, Ismail AS. Accurate and dynamic predictive model for better prediction in medicine and healthcare. Ir J Med Sci 2018 May;187(2):501-513. [doi: 10.1007/s11845-017-1655-3] [Medline: 28756541]
27. Kim E, Caraballo PJ, Castro MR, Pieczkiewicz DS, Simon GJ. Towards more Accessible Precision Medicine: Building a more Transferable Machine Learning Model to Support Prognostic Decisions for Micro- and Macrovascular Complications of Type 2 Diabetes Mellitus. J Med Syst 2019 May 17;43(7):185. [doi: 10.1007/s10916-019-1321-6] [Medline: 31098679]
28. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. N Engl J Med 2017 Jun 29;376(26):2507-2509 [FREE Full text] [doi: 10.1056/NEJMp1702071] [Medline: 28657867]
29. Peterson ED. Machine Learning, Predictive Analytics, and Clinical Practice: Can the Past Inform the Present? JAMA 2019 Nov 22:2283-2284. [doi: 10.1001/jama.2019.17831] [Medline: 31755902]
30. Xiong X, Zhang R, Bi Y, Zhou W, Yu Y, Zhu D. Machine Learning Models in Type 2 Diabetes Risk Prediction: Results from a Cross-sectional Retrospective Study in Chinese Adults. Curr Med Sci 2019 Aug;39(4):582-588. [doi: 10.1007/s11596-019-2077-4] [Medline: 31346994]

31. Fagherazzi G, Ravaud P. Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
32. Das SK. Integrating transcriptome and epigenome: putting together the pieces of the type 2 diabetes pathogenesis puzzle. Diabetes 2014 Sep;63(9):2901-2903 [FREE Full text] [doi: 10.2337/db14-0757] [Medline: 25146470]
33. Franks PW, Pearson E, Florez JC. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. Diabetes Care 2013 May;36(5):1413-1421 [FREE Full text] [doi: 10.2337/dc12-2211] [Medline: 23613601]
34. Aggarwal M, Bozkurt B, Panjrath G, Aggarwal B, Ostfeld RJ, Barnard ND, American College of Cardiology's NutritionLifestyle Committee of the Prevention of Cardiovascular Disease Council. Lifestyle Modifications for Preventing and Treating Heart Failure. J Am Coll Cardiol 2018 Nov 06;72(19):2391-2405 [FREE Full text] [doi: 10.1016/j.jacc.2018.08.2160] [Medline: 30384895]
35. Cardoso R, Nasir K. Primary prevention of cardiovascular disease: 2019 and beyond. Nat Rev Cardiol 2019 Jul;16(7):387-388. [doi: 10.1038/s41569-019-0213-2] [Medline: 31110263]
36. Wilson LF, Baade PD, Green AC, Jordan SJ, Kendall BJ, Neale RE, et al. The impact of changing the prevalence of overweight/obesity and physical inactivity in Australia: An estimate of the proportion of potentially avoidable cancers 2013-2037. Int J Cancer 2019 May 01;144(9):2088-2098 [FREE Full text] [doi: 10.1002/ijc.31943] [Medline: 30357816]
37. Whaley CM, Bollyky JB, Lu W, Painter S, Schneider J, Zhao Z, et al. Reduced medical spending associated with increased use of a remote diabetes management program and lower mean blood glucose values. J Med Econ 2019 Sep;22(9):869-877. [doi: 10.1080/13696998.2019.1609483] [Medline: 31012392]
38. European Medical Information Framework. EMIF URL: http://www.emif.eu/ [accessed 2019-10-21]
39. Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors. Big Data 2015 Dec;3(4):277-287. [doi: 10.1089/big.2015.0020] [Medline: 27441408]
40. Maguire J, Dhar V. Comparative effectiveness for oral anti-diabetic treatments among newly diagnosed type 2 diabetics: data-driven predictive analytics in healthcare. Health Systems 2017 Dec 19;2(2):73-92 [FREE Full text] [doi: 10.1057/hs.2012.20]
41. Kagawa R, Kawazoe Y, Ida Y, Shinohara E, Tanaka K, Imai T, et al. Development of Type 2 Diabetes Mellitus Phenotyping Framework Using Expert Knowledge and Machine Learning Approach. J Diabetes Sci Technol 2017 Jul;11(4):791-799 [FREE Full text] [doi: 10.1177/1932296816681584] [Medline: 27932531]
42. Shankaracharya, Odedra D, Samanta S, Vidyarthi AS. Computational intelligence-based diagnosis tool for the detection of prediabetes and type 2 diabetes in India. Rev Diabet Stud 2012;9(1):55-62 [FREE Full text] [doi: 10.1900/RDS.2012.9.55] [Medline: 22972445]
43. Olivera AR, Roesler V, Iochpe C, Schmidt MI, Vigo A, Barreto SM, et al. Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes - ELSA-Brasil: accuracy study. Sao Paulo Med J 2017;135(3):234-246 [FREE Full text] [doi: 10.1590/1516-3180.2016.0309010217] [Medline: 28746659]
44. Renard LM, Bocquet V, Vidal-Trecan G, Lair M, Couffignal S, Blum-Boisgard C. An algorithm to identify patients with treated type 2 diabetes using medico-administrative data. BMC Med Inform Decis Mak 2011 Apr 14;11:23 [FREE Full text] [doi: 10.1186/1472-6947-11-23] [Medline: 21492480]
45. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
46. Degli Esposti L, Saragoni S, Buda S, Sturani A, Degli Esposti E. Glycemic control and diabetes-related health care costs in type 2 diabetes; retrospective analysis based on clinical and administrative databases. Clinicoecon Outcomes Res 2013;5:193-201 [FREE Full text] [doi: 10.2147/CEOR.S41846] [Medline: 23696709]
47. Musacchio N, Candido R, Cimino A, De Micheli A, Giancaterini A, Monge L, et al. Diabetologist's core competence curriculum: A position statement of the amd (Italian association of medical diabetologists). Diabetes Management 2019;9(4):87-95 [FREE Full text]
48. Liaw ST, Taggart J, Yu H, de Lusignan S, Kuziemsky C, Hayen A. Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers. J Biomed Inform 2014 Dec;52:364-372 [FREE Full text] [doi: 10.1016/j.jbi.2014.07.016] [Medline: 25089026]
49. Gnavi R, Karaghiosoff L, Costa G, Merletti F, Bruno G. Socio-economic differences in the prevalence of diabetes in Italy: the population-based Turin study. Nutr Metab Cardiovasc Dis 2008 Dec;18(10):678-682. [doi: 10.1016/j.numecd.2007.11.004] [Medline: 18395428]
50. Verlato G, Muggeo M, Bonora E, Corbellini M, Bressan F, de Marco R. Attending the diabetes center is associated with increased 5-year survival probability of diabetic patients: the Verona Diabetes Study. Diabetes Care 1996 Mar;19(3):211-213. [doi: 10.2337/diacare.19.3.211] [Medline: 8742563]
51. Bruno G, Merletti F, Biggeri A, Bargero G, Ferrero S, Pagano G, Casale Monferrato Study. Fibrinogen and AER are major independent predictors of 11-year cardiovascular mortality in type 2 diabetes: the Casale Monferrato Study. Diabetologia 2005 Mar;48(3):427-434. [doi: 10.1007/s00125-004-1667-1] [Medline: 15696295]
52. Associazione Medici Diabetologi - AMD. Annali AMD URL: https://aemmedi.it/annali-amd/ [accessed 2019-10-21]

53. Rossi MCE, Nicolucci A, Arcangeli A, Cimino A, De Bigontina G, Giorda C, Associazione Medici Diabetologi Annals Study Group. Baseline quality-of-care data from a quality-improvement program implemented by a network of diabetes outpatient clinics. Diabetes Care 2008 Nov;31(11):2166-2168 [FREE Full text] [doi: 10.2337/dc08-0469] [Medline: 18694979]
54. Rossi MC, Candido R, Ceriello A, Cimino A, Di Bartolo P, Giorda C, et al. Trends over 8 years in quality of diabetes care: results of the AMD Annals continuous quality improvement initiative. Acta Diabetol 2015 Jun;52(3):557-571. [doi: 10.1007/s00592-014-0688-6] [Medline: 25528003]
55. Giorda CB, Nicolucci A, Pellegrini F, Kristiansen CK, Hunt B, Valentine WJ, et al. Improving quality of care in people with Type 2 diabetes through the Associazione Medici Diabetologi-annals initiative: a long-term cost-effectiveness analysis. Diabet Med 2014 May;31(5):615-623. [doi: 10.1111/dme.12366] [Medline: 24246087]
56. Rossi MC, Cristofaro MR, Gentile S, Lucisano G, Manicardi V, Mulas MF, AMD Annals Study Group. Sex disparities in the quality of diabetes care: biological and cultural factors may play a different role for different outcomes: a cross-sectional observational study from the AMD Annals initiative. Diabetes Care 2013 Oct;36(10):3162-3168 [FREE Full text] [doi: 10.2337/dc13-0184] [Medline: 23835692]
57. Viazzi F, Russo GT, Ceriello A, Fioretto P, Giorda C, De Cosmo S, et al. Natural history and risk factors for diabetic kidney disease in patients with T2D: lessons from the AMD-annals. J Nephrol 2019 Aug;32(4):517-525. [doi: 10.1007/s40620-018-00561-3] [Medline: 30478509]
58. Musacchio N, Zilich R, Candido R, Cimino A, De MA, Giancaterini A, et al. Core Competence Curriculum del diabetologo: Position Statement di AMD JAMD 20(2):S15-S32, 2017. Diabetes Management 2019;9(4):87-95 [FREE Full text]
59. Russel SJ, Norvig P. Artificial Intelligence: A Modern Approach. New Jersey 07632: Prentice Hall, Englewood Cliffs; 1995.

Abbreviations

AI: artificial intelligence
AMD: Association of Medical Diabetologists
CCM: chronic care model
DIA&INT: diabetes intelligence
HbA1c: hemoglobin A1c
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 05.11.19; peer-reviewed by F Andreozzi, S De Cosmo, R Kagawa; comments to author 06.01.20; revised version received 09.03.20; accepted 12.04.20; published 22.06.20.

Please cite as:
Musacchio N, Giancaterini A, Guaita G, Ozzello A, Pellegrini MA, Ponzani P, Russo GT, Zilich R, de Micheli A
Artificial Intelligence and Big Data in Diabetes Care: A Position Statement of the Italian Association of Medical Diabetologists
J Med Internet Res 2020;22(6):e16922
URL: http://www.jmir.org/2020/6/e16922/
doi:10.2196/16922
PMID:32568088

Viewpoint

The Adaptive Behavioral Components (ABC) Model for Planning Longitudinal Behavioral Technology-Based Health Interventions: A Theoretical Framework

Sean D Young[1,2], MS, PhD

[1]Institute for Prediction Technology, Department of Informatics, University of California, Irvine, Irvine, CA, United States
[2]Department of Emergency Medicine, UCI School of Medicine, Irvine, CA, United States

Corresponding Author:
Sean D Young, MS, PhD
Institute for Prediction Technology
Department of Informatics
University of California, Irvine
Bren School of Information and Computer Science
Irvine, CA, 92617-4022
United States
Phone: 1 650 223 5448
Email: syoung5@hs.uci.edu

Abstract

A growing number of interventions incorporate digital and social technologies (eg, social media, mobile phone apps, and wearable devices) into their design for behavior change. However, because of a number of factors, including changing trends in the use of technology over time, results on the efficacy of these interventions have been mixed. An updated framework is needed to help researchers better plan behavioral technology interventions by anticipating the needed resources and potential changes in trends that may affect interventions over time. Focusing on the domain of health interventions as a use case, we present the Adaptive Behavioral Components (ABC) model for technology-based behavioral interventions. ABC is composed of five components: basic behavior change; intervention, or problem-focused characteristics; population, social, and behavioral characteristics; individual-level and personality characteristics; and technology characteristics. ABC was designed with the goals of (1) guiding high-level development for digital technology–based interventions; (2) helping interventionists consider, plan for, and adapt to potential barriers that may arise during longitudinal interventions; and (3) providing a framework to potentially help increase the consistency of findings among digital technology intervention studies. We describe the planning of an HIV prevention intervention as a case study for how to implement ABC into intervention design. Using the ABC model to plan future interventions might help to improve the design of and adherence to longitudinal behavior change intervention protocols; allow these interventions to adapt, anticipate, and prepare for changes that may arise over time; and help to potentially improve intervention behavior change outcomes. Additional research is needed on the influence of each of ABC's components to help improve intervention design and implementation.

(J Med Internet Res 2020;22(6):e15563) doi:10.2196/15563

KEYWORDS

health behavior; risk behavior; behavioral medicine; public health informatics; consumer health informatics; psychological theory

Introduction

Digital and social technologies (eg, social media, smartphone apps, and wearable devices) have promising potential for achieving rapid and widespread health behavior change. Social media has been used to change and predict a number of health-related behaviors, including HIV testing and sexually transmitted diseases, suicide prevention, car crashes, and opioid-related emergency department visits [1-7]; Fitbits and self-tracking devices have been proposed as intervention tools to increase exercise [8,9] and reduce stress; health systems and insurers have integrated wearable device, social media and patient health/medical data to try to improve clinical outcomes [10,11]; and smartphone apps have be studied for their potential to improve a variety of health behaviors and outcomes, such as weight loss and diabetes self-management [12-14].

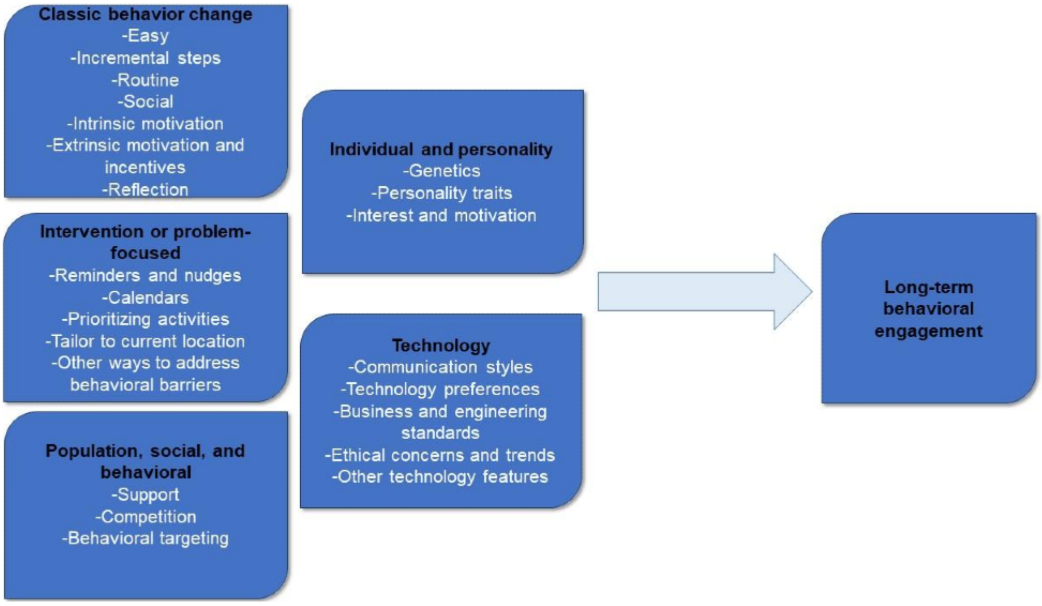
However, results have been mixed on whether and how digital technologies might change people's health behaviors [15-20]. Although some of these differences may be related to common methodological and study assessment–related reasons (eg, differences in study duration or outcomes), there are also a number of potential intervention design–related reasons for these inconsistences. For example, a large volume of behavioral psychology research suggests that small contextual changes have a dramatic effect on behavior [21-24], such as the size of a button on a website, or the way in which information is communicated and displayed. Although contextual issues can and have been found to impact intervention success regardless of whether it is an online or offline study (eg, age, race/ethnicity, and sex of experimenter may affect participant adherence and engagement), these issues become more complicated when delivering interventions digitally because of the frequency and ease of changing variables in digital technologies compared with offline interventions (eg, placement of buttons and text, inclusion of social interaction, and gamification), as well as reliance on technology companies (eg, Facebook and Twitter) to maintain similar and stable versions of their products. These issues have become increasingly important with the current COVID-19 pandemic, as there is growing need to integrate new technologies into research and clinical care to address COVID-19 policies, such as social distancing and stay-at-home orders.

Due to changing trends in design and use of digital technologies, it is, therefore, extremely difficult to exactly replicate a technology-based intervention study, as the earlier study might have occurred on a now outdated software platform. This creates an additional problem as grant applications often require planning 3- to 5-year studies in advance despite the potential technological changes that may occur during that period. How can researchers designing digital technology–based interventions address, or at least anticipate, these issues? New theories are needed to build on existing health informatics and behavioral technology-based intervention models [25].

This manuscript proposes a theoretical framework, called the Adaptive Behavioral Components (ABC) theory for technology-based interventions, with the goals of (1) guiding high-level development for digital health technology–based interventions; (2) helping interventionists consider, plan for, and adapt to potential barriers that may arise during longitudinal interventions; and (3) providing a framework to potentially help increase consistency of findings among digital technology intervention studies. We seek to describe and synthesize categories of prior research (below) into a new theory to help researchers design digital health technology–based interventions.

ABC incorporates research across the fields of social and behavioral psychology, informatics, and marketing to develop a model tailored to the needs of digital technology interventionists. The model is based on five overarching factors needed for guiding a sustainable technology-based intervention: (1) basic behavior change components; (2) intervention and problem-focused characteristics; (3) population, social, and behavioral characteristics; (4) individual-level and personality characteristics; and (5) technology characteristics (Figure 1).

Figure 1. Adaptive Behavioral Components Theory (ABC) for Technology-based Interventions. By helping researchers to consider these five high-level components, the model provides guidance on how to plan for the needs and potential changes during a technology-based health intervention.



A large body of research, including theoretical modeling, has been conducted on many of these five areas. ABC is not meant to replace existing theories specific to one or more of these components, but rather to provide a guide for when and how to think about the inter-related, diverse, and overarching concepts affecting behavior change that are often not considered within behavioral interventions. The first four components of ABC can apply broadly to behavioral interventions, whereas the last component is specific to digital technology–based interventions. Although the model might also be applied outside of health, this manuscript focuses on health applications to narrow the scope of discussion and provide context. The five components of ABC are described in more detail below, along with a case study of how it has been learned and might be applied in research settings.

The ABC Framework

Basic Behavior Change Components

There are a number of behavior change methods and theories, such as social normative and social cognitive theory–based interventions, that have been broadly and successfully applied to change health behaviors, including HIV risk behaviors [26] and smoking cessation [27]. These methods are typically rooted in social and behavioral psychological theory [20] and broadly suggest that interventions will be more efficacious if they (1) make it easy to engage in the target behavior [28-32], (2) make the behavior routine rather than infrequent and unpredictable [33,34], (3) incorporate social components such as support and/or competition [2,35-37], (4) build change incrementally (eg, dieters who focused on the daily steps of eating healthy were more likely to adhere to their healthy eating routine compared with those who focused on the end goal of losing weight [and these individuals actually gained more weight at follow-up]) [38], (5) develop intrinsic [39-41] and (6) extrinsic motivation (being cautious about the sometimes negative effects of incentives on behavior) [42,43], and (7) provide ways for people to reflect on their progress [44,45]. Integration of these basic behavior change components may help to improve intervention behavior change–related outcomes.

Intervention/Problem-Focused Characteristics

Interventions may benefit from further tailoring to address the *specific* behavioral barriers preventing people from conducting the intended behavior. These intervention or problem-focused characteristics are typically contextual factors that affect people’s willingness and ability to participate and remain engaged in an intervention. For example, people might not engage in behaviors because of a variety of reasons, including lack of remembering, interest, or time; belief that the intervention is unlikely to help or may harm them; cost; or that the intervention is ill-suited to their lifestyle, values, or expectations. The specific intervention framework should be guided by these needs [46]. For example, an intervention designed to send mobile-delivered medication reminders would likely be more beneficial to individuals who are forgetting to take their medication compared with individuals who fear the side effects of the medication. Similarly, environmental and immediate contextual-level tailoring (eg, just-in-time interventions) may benefit individuals who need to be reminded to engage in activities at certain times or in certain locations [47]. For example, a text message reminder can be sent as a reminder tool to attempt to prevent drug use for a substance-using individual who is walking in an area detected by the phone to be a high-risk area for substance use, an example of timecasting [48]. By understanding the reasons why people are not engaging in a specific behavior, intervention characteristics can be tailored to optimize addressing that barrier to behavior change.

Population/Social/Behavioral Characteristics

Tailoring interventions to people’s differences in population, social, and behavioral characteristics may help to further improve intervention efficacy. For example, tailoring intervention content to the target population (eg, creating racially and culturally tailored online recruitment advertisements for a study among African Americans) may increase recruitment efficiency [49]. In a study demonstrating this approach, African American men had increased odds of clicking on a Web-based advertisement that included an image of an African American model vs a white model [50]. Similarly, tailoring interventions to include individuals who share similar demographic characteristics and preferences (ie, peers) may be more efficacious compared with interventions that do not tailor for these factors [51]. Tailoring interventions to an individual’s behavioral state might also lead to higher engagement compared with not including this tailoring. For example, online smoking cessation health information may have higher click-through and engagement rates if delivered to people who have been searching on Google for ways to quit smoking compared with individuals browsing on Facebook to talk with friends.

Individual-Level/Personality Characteristics

Despite the success of social- and community-based interventions, many people within these interventions do not follow the group norm because of individual-level differences, such as differences in experience, genetics, personality, and disposition [52]. For example, in a study on factors affecting health, the personality factor, that is, conscientiousness, predicted all study outcomes, with individuals identified as being less responsible and less self-controlled tending to report poorer health, be more overweight, and engage in more substance use. Tailoring interventions at these individual levels, such as by targeting familial, genetic, and/or personality characteristics, might increase intervention efficacy. For example, participants who have personality traits identifying them as “more social” might be targeted as peer role models for peer-led interventions [53-55]. Similarly, individuals with genetic predisposition for alcohol use disorder [56] might be targeted for an alcohol-related intervention to help researchers learn more about genetic factors/predisposition and improve targeting of at-risk individuals. However, research on the role and methods of developing interventions based on individual-level/personality characteristics has been limited, leading researchers to call for additional research on this topic [57].

Technology Characteristics

Although the above four factors affect and should be considered when conducting any type of behavior change intervention, the technology characteristics component is specific and critical to the success of *digital technology* interventions. It is also perhaps the component that requires the most planning of all the five factors, because of the changing needs and trends in technology use. Intervention engagement can be affected by a variety of technology characteristics, such as current trends in popular communication style (eg, use of video as a preferred communication medium vs text), changes in features of a technology (eg, adding an online community feature to a software app may change user engagement rates), and changes in ethical concerns (eg, social media engagement may decline after a recent security breach on that social media platform) [58]. These factors might have a large impact on intervention efficacy. Understanding and preparing for technology characteristics and potential adaptations is, therefore, essential for the delivery of effective longitudinal technology-based interventions.

Discussion

Lessons Learned and How to Apply

As a case study identifying the five components of ABC and how to apply them, we will describe the process for planning an intervention to increase the initiation of pre-exposure prophylaxsis (PrEP) as a method of HIV prevention among African American and Latino men who have sex with men (MSM) who are aged 18 years or older and living in Los Angeles. The primary behavior we wish to change is, therefore, to increase the number of participants initiating PrEP. Table 1 can be used as a reference for when and how the ABC framework can be incorporated into planning this intervention.

Table 1. Applying the Adaptive Behavioral Components (ABC) model for technology-based interventions. This table provides an example of how a researcher might apply the ABC framework in an intervention designed to increase uptake of pre-exposure prophylaxis among African American and Latino men who have sex with men at risk for HIV. The examples below are not exhaustive but are meant to assist readers in understanding how to integrate ABC into intervention design.

Category and subcategory	Example
Classic behavior change	
Easy	Provide online diagnosis and prescription referral as an option to increase PrEP[a] accessibility
Incremental steps	Gradually discuss and promote PrEP with participants, beginning with friendly topics first to gain trust
Routine	Routine (eg, weekly) communications with study participants
Social	Encourage discussions and social interaction among participants
Intrinsic motivation	Educate participants on the importance of prevention and safe sex
Extrinsic motivation	Provide products/services that have tangible value to participants (eg, free PrEP and HIV self-test)
Reflection	Encourage participants to talk about their experiences using PrEP so that they reflect on and think about it
Intervention/problem-focused	
Reminders/nudges	Remind participants to make an appointment and use PrEP
Calendars	Codevelop a calendar with them with dates for when they will schedule an appointment
Prioritizing activities	Cocreate a list of priorities with them to help them fit a PrEP medical visit and use into their priorities
Tailor to current location	Discuss PrEP when participants are available based on their location and schedule
Population/social/behavioral	
Social support	Develop a peer support network where peer role models deliver the intervention
Competition	Integrate friendly games
Behavioral targeting	Provide recruitment advertisements and/or HIV prevention information to participants at the right time and context (eg, when participants are requesting or searching for information about safe sex)
Individual/personality	
Genetics	Tailor the intervention based on genetic differences in response found in research
Personality traits	Tailor the intervention based on personality differences in response found in research
Interest/motivation	Tailor the intervention based on interest/motivation (eg, individuals who are less interested may need greater financial incentives)
Technology	
Communication styles	Offer multiple methods of communication (eg, chat, video, and text message)
Technology preferences	Deliver the intervention on platforms that are already being used by minority men who have sex with men (ie, the target population)
Business/engineering standards	Awareness and planning based on current and potential future changing trends in the technology platforms being used
Ethical considerations/trends	Awareness and planning technology features that address current and potential future ethical considerations (eg, for an HIV prevention intervention, private communication is important to reduce experience of stigma)
Other technology features	Deliver the intervention on multiple technology platforms to avoid risks from one platform

[a] PrEP: pre-exposure prophylaxsis.

Although there are numerous reasons why individuals do not take PrEP, using the ABC framework to analyze the intervention/problem-focused characteristics teaches us that we need to narrow to focus on only one of these reasons (stigma) and develop an intervention designed to decrease stigma to increase PrEP use. As stigma is a social construct (created and perpetuated by one's peers) [59], we seek to counteract stigma that prevents people from using PrEP by creating an intervention that will leverage peers to promote PrEP uptake (ie, we will deliver a peer-led, community-based intervention). Within the intervention framework, we will attempt to include many of the basic behavior change components. For example, we will make it easy for participants to request PrEP by allowing them to receive an online tele-consultation and request PrEP online, making it easier (and less stigmatizing) than typical methods of having to visit and get tested at in-person clinics. We will also hire peer leaders to gradually become friends and then promote PrEP behavior change with participants each week, incentivizing routine discussions with the peer leaders.

We will integrate the population/social/behavioral characteristics of ABC by designing recruitment materials, intervention content, and processes that are tailored for this population (eg, recruitment advertisements include sexually provocative images of minority MSM; advertisements will be placed on websites where minority MSM are already searching for safe sex information [eg, Google ad words]; and HIV prevention conversations will be delivered by minority MSM peer leaders and focused on issues relevant to minority MSM).

Although the intervention is designed for African American and Latino MSM, we anticipate individual differences such that not all individuals will respond to the intervention as planned. To address this, we will meet peer leaders each week to learn about the barriers and facilitators in promoting PrEP uptake among each individual participant, to refine and further tailor the intervention to individuals who are not responsive to the intervention (eg, some participants may require more personal time/education than the average group member).

Finally, we wish to use technologies to scale the peer-led PrEP uptake intervention because of a technology-based intervention's potential for rapidly and cost-effectively scaling behavior change. We seek to use social media and mobile social technologies to deliver the intervention because of their alignment with and ability to meet the requirements stated above (eg, social media technologies allow African American and Latino MSM peer community leaders to communicate with and spread PrEP uptake–related behavior change information to other African American and Latino MSM both within a [private] group and/or individually). The Harnessing Online Peer Education (HOPE) intervention, a peer-led online community-based behavior change intervention, is one example of this type of intervention and one that we seek to use. Although HOPE interventions have been focused on different populations and behavioral outcomes, including HIV testing (HOPE HIV), mental health, and opioid misuse, with efficacy demonstrated in changing primary behavioral outcomes [2,51,60,61], all members of this family of intervention rely on using social technologies as a method for delivering a social normative (eg, peer-based) intervention. However, deciding the specific technology component(s) to be used in HOPE is perhaps the most important and difficult aspect in our study plan because of the changing trends in technologies.

For example, a multiwave, 5-year HOPE HIV testing intervention was funded in 2014 based on earlier methods and results from HOPE interventions conducted in 2010 and 2012 that showed success in using Facebook groups as an intervention delivery platform [2,51]. However, from 2010 to 2014, and throughout the 5 years following 2014, a large number of changes have occurred on social media, including Facebook no longer being the sole dominant technology used for online communities or forums among all age groups and populations. With the introduction of other platforms, such as Snapchat and Instagram, use of Facebook has waned, especially among youth [62], which might reduce the engagement and potential efficacy of interventions delivered on Facebook that rely on previous rates of engagement. Similarly, throughout a 5-year period, Facebook has made a number of changes to their interface, affecting HOPE intervention engagement.

For example, in 2018, Facebook's Mark Zuckerberg made a public statement that Facebook would be changing the way their algorithm displayed content, upsetting many individuals and businesses who had relied on previous methods for posting on Facebook that might no longer work [63]. Previous longitudinal behavioral interventions on Facebook were impacted by these changes. For example, in an HIV prevention intervention run by our team using Facebook groups during this period, we have been finding that Facebook platform changes have resulted in intervention participants receiving fewer notifications about HIV prevention and testing compared with control (Facebook) group participants. Although the intervention group remains significantly more engaged in posting compared with the control group (as intended by the intervention), the large number of posts within the intervention group combined with the changes to the way Facebook groups and the Facebook algorithm are used initially resulted in fewer testing-related posts being viewed by the intervention group compared with the control group. As a result, our team has learned the importance of delivering technology-based interventions through multiple technology platforms to minimize the impact on the intervention because of changes that may occur on one of the technology platforms or by changing trends in the types of technologies used by participants.

We responded to these experiences by learning that we need to modify the way technology-based interventions are designed and delivered. In the future, we seek to develop a HOPE intervention that leverages multiple platforms, including Facebook, Instagram, Snapchat, and YouTube (or whichever platforms that are dominant at that time), and ensures that the elements of ABC can be integrated into one or more of these platforms. By delivering the intervention across multiple platforms, we seek to (1) reduce the risk that changes in one or more of the technology platforms alters intervention delivery; (2) reduce the risk that major changes to the study protocol might be needed at a later period; (3) allow flexibility within the study protocol for it to be tailored to incorporate additional social technologies in the future (eg, by leveraging multiple technologies to deliver the intervention, each of these play a smaller role and allows for technologies to be replaced over time as needed with trends while keeping the scope of work intact); (4) allow the ability to simultaneously tailor the intervention to different subdemographic groups and populations (ie, tailoring for age, race, and ethnicity based on the social technologies used most frequently by those groups); and (5) allow the ability to incorporate multimedia methods and trends in communication, such as use of social technologies that are designed for sharing of text, images, videos, and/or other methods such as virtual reality.

We also recommend to our own team and other researchers to track every available relevant digital metric (eg, click-through rates, downloads, and likes) and to conduct a pilot study before proceeding with an intervention. These digital data can be very useful in informing potential unknown problems that might occur during an intervention [64]. They are also important as they can help monitor and address unforeseen circumstances that may occur in the middle of a digital intervention (eg, a prolonged drop in engagement on a digital platform may be a signal for the need to shift the intervention methods/content to a different digital platform). Taken together, we have learned the importance of both accounting for changes in technology use and the need to anticipate and adapt various behavior change components across technologies to improve the efficacy and durability of digital behavioral interventions.

Conclusions

The ABC model was designed as a guide both to assist in planning digitally delivered behavior change interventions and to help anticipate the needed resources and potential changes in trends that may affect an intervention over time. Using the ABC model to plan future interventions may help to improve the design of and adherence to longitudinal behavior change intervention protocols; allow these interventions to adapt, anticipate, and prepare for changes that may arise over time; and help to potentially improve intervention behavior change outcomes.

Acknowledgments

The authors would like to thank Hendricks Brown for feedback on previous versions of this manuscript. The authors wish to thank the National Institute on Mental Health (5R01MH106415), National Institute on Drug Abuse (C-PIM grant), National Human Genome Research Institute (U01), and National Institute for Allergy and Infectious Diseases (7R01AI132030) for funding this work. This manuscript does not reflect the opinions or views of the funders.

Authors' Contributions

SDY developed the theory and wrote the manuscript.

Conflicts of Interest

SDY receives royalties from Harper Collins Publishers (US edition), Penguin (UK edition), and other publishers for the book, Stick with It. He is also an advisor/consultant (receiving compensation) to companies and startups related to behavior change.

References

1. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In: Proceedings of the Seventh International Conference on Weblogs and Social Media. 2013 Presented at: ICWSM'13; July 8–11, 2013; Cambridge, Massachusetts, USA URL: https://www.microsoft.com/en-us/research/publication/predicting-depression-via-social-media/
2. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. Ann Intern Med 2013 Sep 3;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
3. Young SD, Zheng K, Chu LF, Humphreys K. Internet searches for opioids predict future emergency department heroin admissions. Drug Alcohol Depend 2018 Sep 1;190:166-169 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.05.009] [Medline: 30036853]
4. Young SD, Lee S, Perez H, Gill N, Gelberg L, Heinzerling K. Social media as an emerging tool for reducing prescription opioid misuse risk factors. Heliyon 2020 Mar;6(3):e03471. [doi: 10.1016/j.heliyon.2020.e03471]
5. Young SD, Wang W, Chakravarthy B. Crowdsourced traffic data as an emerging tool to monitor car crashes. JAMA Surg 2019 May 22;154(8):777-778 [Epub ahead of print]. [doi: 10.1001/jamasurg.2019.1167] [Medline: 31116356]
6. Young SD, Mercer N, Weiss RE, Torrone EA, Aral SO. Using social media as a tool to predict syphilis. Prev Med 2018 Apr;109:58-61 [FREE Full text] [doi: 10.1016/j.ypmed.2017.12.016] [Medline: 29278678]
7. Garett R, Liu S, Young SD. The relationship between social media use and sleep quality among undergraduate students. Inf Commun Soc 2018;21(2):163-173 [FREE Full text] [doi: 10.1080/1369118X.2016.1266374] [Medline: 29628784]
8. Hartman SJ, Nelson SH, Weiner LS. Patterns of Fitbit use and activity levels throughout a physical activity intervention: exploratory analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2018 Feb 5;6(2):e29 [FREE Full text] [doi: 10.2196/mhealth.8503] [Medline: 29402761]
9. Jiang JY, Chao Z, Bertozzi AL, Wang W, Young SD, Needell D. Learning to Predict Human Stress Level with Incomplete Sensor Data from Wearable Devices. In: Proceedings of the 28th ACM International Conference on Information and

Knowledge Management. 2019 Presented at: CIKM'19; November 3 - 7, 2019; Beijing, China p. 2773-2781. [doi: 10.1145/3357384.3357831]
10. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
11. Young SD. Social media as a new vital sign: commentary. J Med Internet Res 2018 Apr 30;20(4):e161 [FREE Full text] [doi: 10.2196/jmir.8563] [Medline: 29712631]
12. Adu MD, Malabu UH, Malau-Aduli AE, Malau-Aduli BS. Users' preferences and design recommendations to promote engagements with mobile apps for diabetes self-management: Multi-national perspectives. PLoS One 2018;13(12):e0208942 [FREE Full text] [doi: 10.1371/journal.pone.0208942] [Medline: 30532235]
13. An R, Ji M, Zhang S. Effectiveness of social media-based interventions on weight-related behaviors and body weight status: review and meta-analysis. Am J Health Behav 2017 Nov 1;41(6):670-682. [doi: 10.5993/AJHB.41.6.1] [Medline: 29025495]
14. Wharton CM, Johnston CS, Cunningham BK, Sterner D. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J Nutr Educ Behav 2014;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]
15. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 3;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
16. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: scoping review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
17. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
18. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
19. Tong HL, Laranjo L. The use of social features in mobile health interventions to promote physical activity: a systematic review. NPJ Digit Med 2018;1:43 [FREE Full text] [doi: 10.1038/s41746-018-0051-3] [Medline: 31304323]
20. Young SD. Stick With It: A Scientifically Proven Process For Changing Your Life-for Good. New York: Harper; 2017.
21. Cialdini RB. Influence: The Psychology Of Persuasion. Revised Edition. New York: Quill; 2006.
22. Gladwell M. The Tipping Point: How Little Things Can Make a Big Difference. Boston: Back Bay Books; 2002.
23. Miller DT, Prentice DA. Psychological levers of behavior change. In: Shafir E, editor. The Behavioral Foundations of Public Policy. Princeton: Princeton University Press; 2013:301-309.
24. Ross L, Nisbett RE. The Person and the Situation: Perspectives of Social Psychology. England: Mcgraw-Hill; 1991.
25. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
26. Latkin CA, Sherman S, Knowlton A. HIV prevention among drug users: outcome of a network-oriented peer outreach intervention. Health Psychol 2003 Jul;22(4):332-339. [doi: 10.1037/0278-6133.22.4.332] [Medline: 12940388]
27. Ford P, Clifford A, Gussy K, Gartner C. A systematic review of peer-support programs for smoking cessation in disadvantaged groups. Int J Environ Res Public Health 2013 Oct 28;10(11):5507-5522 [FREE Full text] [doi: 10.3390/ijerph10115507] [Medline: 24169412]
28. Iyengar SS, Lepper MR. When choice is demotivating: can one desire too much of a good thing? J Pers Soc Psychol 2000 Dec;79(6):995-1006. [doi: 10.1037//0022-3514.79.6.995] [Medline: 11138768]
29. D’Angelo JD, Toma C. There are plenty of fish in the sea: the effects of choice overload and reversibility on online daters’ satisfaction with selected partners. Media Psychol 2016;20(1):1-27 [FREE Full text] [doi: 10.1080/15213269.2015.1121827]
30. Campbell CA, Hahn RA, Elder R, Brewer R, Chattopadhyay S, Fielding J, Task Force on Community Preventive Services. The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. Am J Prev Med 2009 Dec;37(6):556-569. [doi: 10.1016/j.amepre.2009.09.028] [Medline: 19944925]
31. de Vries H, Eggers SM, Bolman C. The role of action planning and plan enactment for smoking cessation. BMC Public Health 2013 Apr 26;13:393 [FREE Full text] [doi: 10.1186/1471-2458-13-393] [Medline: 23622256]
32. Young SD, Monin B, Owens D. Opt-out testing for stigmatized diseases: a social psychological approach to understanding the potential effect of recommendations for routine HIV testing. Health Psychol 2009 Nov;28(6):675-681 [FREE Full text] [doi: 10.1037/a0016395] [Medline: 19916635]
33. Mullan B, Henderson J, Kothe E, Allom V, Orbell S, Hamilton K. The role of habit and perceived control on health behavior among pregnant women. Am J Health Behav 2016 May;40(3):291-301. [doi: 10.5993/AJHB.40.3.1] [Medline: 27103408]
34. Wood W, Rünger D. Psychology of habit. Annu Rev Psychol 2016;67:289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]

35. Young SD, Nussbaum AD, Monin B. Potential moral stigma and reactions to sexually transmitted diseases: evidence for a disjunction fallacy. Pers Soc Psychol Bull 2007 Jun;33(6):789-799. [doi: 10.1177/0146167207301027] [Medline: 17488871]
36. Schweier R, Romppel M, Richter C, Hoberg E, Hahmann H, Scherwinski I, et al. A web-based peer-modeling intervention aimed at lifestyle changes in patients with coronary heart disease and chronic back pain: sequential controlled trial. J Med Internet Res 2014 Jul 23;16(7):e177 [FREE Full text] [doi: 10.2196/jmir.3434] [Medline: 25057119]
37. Nishiyama Y, Okoshi T, Yonezawa T, Nakazawa J, Takashio K, Tokuda H. Toward health exercise behavior change for teams using Lifelog sharing models. IEEE J Biomed Health Inform 2016;20(3):775-786. [doi: 10.1109/jbhi.2015.2478903]
38. Freund AM, Hennecke M. Changing eating behaviour vs losing weight: the role of goal focus for weight loss in overweight women. Psychol Health 2012 Oct;27(Suppl 2):25-42. [doi: 10.1080/08870446.2011.570867] [Medline: 21678179]
39. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. JMIR Serious Games 2019 Feb 13;7(1):e11444 [FREE Full text] [doi: 10.2196/11444] [Medline: 30758293]
40. Lemos MS, Veríssimo L. The Relationships between Intrinsic Motivation, Extrinsic Motivation, and Achievement, Along Elementary School. Procedia Soc Behav Sci 2014;112:930-938. [doi: 10.1016/j.sbspro.2014.01.1251]
41. Ng B. The neuroscience of growth mindset and intrinsic motivation. Brain Sci 2018 Jan 26;8(2):pii: E20 [FREE Full text] [doi: 10.3390/brainsci8020020] [Medline: 29373496]
42. Vohs KD, Mead NL, Goode MR. Merely activating the concept of money changes personal and interpersonal behavior. Curr Dir Psychol Sci 2008;17(3):208-212. [doi: 10.1111/j.1467-8721.2008.00576.x]
43. Higgins ST, Petry NM. Contingency management. Incentives for sobriety. Alcohol Res Health 1999;23(2):122-127 [FREE Full text] [Medline: 10890806]
44. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295x.84.2.191]
45. Yee N, Bailenson J. The Proteus effect: the effect of transformed self-representation on behavior. Hum Commun Res 2007 Jul;33(3):271-290. [doi: 10.1111/j.1468-2958.2007.00299.x]
46. Flottorp S, Oxman AD. Identifying barriers and tailoring interventions to improve the management of urinary tract infections and sore throat: a pragmatic study using qualitative methods. BMC Health Serv Res 2003 Feb 4;3(1):3 [FREE Full text] [doi: 10.1186/1472-6963-3-3] [Medline: 12622873]
47. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
48. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas L, Wingood G, et al. A computational future for preventing HIV in minority communities: how advanced technology can improve implementation of effective programs. J Acquir Immune Defic Syndr 2013 Jun 1;63(Suppl 1):S72-S84 [FREE Full text] [doi: 10.1097/QAI.0b013e31829372bd] [Medline: 23673892]
49. Wingood GM, DiClemente RJ. The ADAPT-ITT model: a novel method of adapting evidence-based HIV Interventions. J Acquir Immune Defic Syndr 2008 Mar 1;47(Suppl 1):S40-S46. [doi: 10.1097/QAI.0b013e3181605df1] [Medline: 18301133]
50. Sullivan PS, Khosropour CM, Luisi N, Amsden M, Coggia T, Wingood GM, et al. Bias in online recruitment and retention of racial and ethnic minority men who have sex with men. J Med Internet Res 2011 May 13;13(2):e38 [FREE Full text] [doi: 10.2196/jmir.1797] [Medline: 21571632]
51. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
52. Huang I, Lee JL, Ketheeswaran P, Jones CM, Revicki DA, Wu AW. Does personality affect health-related quality of life? A systematic review. PLoS One 2017;12(3):e0173806 [FREE Full text] [doi: 10.1371/journal.pone.0173806] [Medline: 28355244]
53. Young SD, Holloway I, Jaganath D, Rice E, Westmoreland D, Coates T. Project HOPE: online social network changes in an HIV prevention randomized controlled trial for African American and Latino men who have sex with men. Am J Public Health 2014 Sep;104(9):1707-1712. [doi: 10.2105/AJPH.2014.301992] [Medline: 25033137]
54. Jaganath D, Gill HK, Cohen AC, Young SD. Harnessing Online Peer Education (HOPE): integrating C-POL and social media to train peer leaders in HIV prevention. AIDS Care 2012;24(5):593-600 [FREE Full text] [doi: 10.1080/09540121.2011.630355] [Medline: 22149081]
55. Young SD, Szekeres G, Coates T. The relationship between online social networking and sexual risk behaviors among men who have sex with men (MSM). PLoS One 2013;8(5):e62271 [FREE Full text] [doi: 10.1371/journal.pone.0062271] [Medline: 23658716]
56. Reilly MT, Noronha A, Goldman D, Koob GF. Genetic studies of alcohol dependence in the context of the addiction cycle. Neuropharmacology 2017 Aug 1;122:3-21 [FREE Full text] [doi: 10.1016/j.neuropharm.2017.01.017] [Medline: 28118990]

57. Mõttus E, Briley DA, Zheng A, Mann FD, Engelhardt LE, Tackett JL, et al. Kids becoming less alike: A behavioral genetic analysis of developmental increases in personality variance from childhood to adolescence. J Pers Soc Psychol 2019 Sep;117(3):635-658. [doi: 10.1037/pspp0000194] [Medline: 30920282]
58. Garett R, Young SD. Health care gamification: a study of game mechanics and elements. Tech Know Learn 2019;24(3):341-353. [doi: 10.1007/s10758-018-9353-4]
59. Goffman E. Stigma: Notes on the Management of Spoiled Identity. New York: Touchstone; 1986.
60. Young SD, Koussa M, Lee S, Perez H, Gill N, Gelberg L, et al. Feasibility of a social media/online community support group intervention among chronic pain patients on opioid therapy. J Addict Dis 2018;37(1-2):96-101 [FREE Full text] [doi: 10.1080/10550887.2018.1557992] [Medline: 30614403]
61. Young SD, Heinzerling K. The Harnessing Online Peer Education (HOPE) intervention for reducing prescription drug abuse: A qualitative study. J Subst Use 2017;22(6):592-596 [FREE Full text] [doi: 10.1080/14659891.2016.1271039] [Medline: 29551953]
62. Susannah Fox. Research URL: https://susannahfox.com/research/ [accessed 2019-01-23]
63. Facebook. 2018 Jan 11. Mark Zuckerberg URL: https://www.facebook.com/zuck/posts/10104413015393571 [accessed 2019-04-08]
64. Young SD, Garett R. Ethical issues in addressing social media posts about suicidal intentions during an online study among youth: case study. JMIR Ment Health 2018 May 3;5(2):e33 [FREE Full text] [doi: 10.2196/mental.8971] [Medline: 29724707]

Abbreviations

ABC: Adaptive Behavioral Components
HOPE: Harnessing Online Peer Education
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis

Edited by G Eysenbach; submitted 26.07.19; peer-reviewed by K Bagot, G Kolostoumpis; comments to author 02.10.19; revised version received 28.11.19; accepted 16.12.19; published 26.06.20.

Please cite as:
Young SD
The Adaptive Behavioral Components (ABC) Model for Planning Longitudinal Behavioral Technology-Based Health Interventions: A Theoretical Framework
J Med Internet Res 2020;22(6):e15563
URL: https://www.jmir.org/2020/6/e15563
doi:10.2196/15563
PMID:32589152

Original Paper

A Digital Pornography Literacy Resource Co-Designed With Vulnerable Young People: Development of "The Gist"

Angela C Davis[1,2], PhD; Cassandra JC Wright[2,3], PhD; Stacey Murphy[4], BDes; Paul Dietze[1,2], PhD; Meredith J Temple-Smith[5], DHSc; Margaret E Hellard[1,2,6,7], PhD; Megan SC Lim[1,2,8], PhD

[1]Burnet Institute, Melbourne, Australia

[2]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia

[3]Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia

[4]Sheda, Melbourne, Australia

[5]Department of General Practice, Melbourne Medical School, University of Melbourne, Melbourne, Australia

[6]Doherty Institute and School of Population and Global Health, University of Melbourne, Melbourne, Australia

[7]Department of Infectious Diseases, The Alfred Hospital, Melbourne, Australia

[8]Melbourne School of Population and Global Health, Department of General Practice, University of Melbourne, Melbourne, Australia

Corresponding Author:
Angela C Davis, PhD
Burnet Institute
GPO Box 2284
Melbourne, 3181
Australia
Phone: 61 3 9282 2111
Email: angela.davis@burnet.edu.au

Abstract

Background: The impact of viewing pornography at a young age on the sexual health of subgroups of young people is an important public health issue. However, the topic is complex and extremely sensitive, and best practices for research and harm reduction are yet to be defined. Drawing on cross-disciplinary approaches, such as co-design, is one way to achieve a better understanding of the issue among vulnerable young people and to create needs-based and evidence-informed digital resources to promote pornography literacy.

Objective: The objective of this study was to co-design a relevant, usable, and acceptable digital prototype to address the pornography literacy needs of vulnerable young people.

Methods: In total, 17 young people aged between 14 and 23 years who were engaged in youth services programs or alternative education programs were recruited to participate in 4 co-design workshops with a multidisciplinary design team.

Results: Although the participants could identify problems with pornography and critique its messages, they lacked the information to understand alternative healthy attitudes and behaviors. A digital resource that provides detailed and practical information about sex, sexual ethics, and relationships may help vulnerable young people to identify and contrast with any problematic messages they receive from both pornography and society. Embedding this information with pornography literacy messages may be a more effective way of addressing underlying attitudes. Acknowledging information-seeking patterns and leveraging user interaction patterns from commonly used digital platforms among users may enhance engagement with resources. Importantly, digital platforms are perceived among this group as a source of anonymous secondary information but would not be organically accessed among this group without face-to-face conversations as an access point.

Conclusions: This paper highlights the potential for pornography literacy to be embedded within real and practical information about having sex, navigating sexuality, and healthy relationships. The study findings include important recommendations for the conceptualization of digital pornography literacy programs and opportunities for cross-disciplinary co-design research to address complex and emerging health issues.

(J Med Internet Res 2020;22(6):e15964) doi:10.2196/15964

KEYWORDS

co-design; pornography literacy; sexual health; sex education

Introduction

Background

Young people's increasing access to free web-based pornography has led to concerns over its impact on their attitudes and behavior [1]. Pornography is defined as sexually explicit pictures, texts, or other material created to cause sexual arousal [2]. However, in reality, web-based pornography includes a wide variety of material and genres that depict diverse sexual scenes [3-5]. The literature suggests that condom use in heterosexual pornography is rare (2%-3%) and that gender inequality in mainstream online videos is common [3,5,6]. For instance, female actors were found to be significantly more likely to be shown being dominated or degraded [3] and were four times less likely than male actors to be shown achieving orgasm [4]. Recent research with 15- to 29-year-old Australians found that 90% had ever viewed pornography, with 81% of young men and 22% of young women reporting at least weekly use [7]. Pornography exposure is associated with sexual risk factors, including decreased age of first sexual experience, more lifetime sexual partners, and not using a condom during the last sexual encounter [8-10]. Exposure has been associated with increased sexual aggression and sexist attitudes [11], and cross-sectional studies report links with relationship problems and depression [12-16]. However, harmful effects have not consistently been reported in research, with some studies showing mixed findings [17]. Given the scale of the issue, it is a public health imperative to better understand how viewing this content frequently and from a young age may impact on sexual scripts, which include the knowledge, beliefs, and attitudes that guide sexual decision making [18,19]. However, the topic is evolving, complex, and extremely sensitive, and best practices for research and harm reduction are yet to be defined. Drawing on cross-disciplinary approaches, such as co-design, is one way to achieve a better understanding of the issue among young people and to create needs-based and evidence-informed responses.

Vulnerable Young People at Risk

Pornography is not homogenous, nor are its audiences [20]. Informed by the social cognitive theory, researchers have suggested that there may be differential effects of viewing pornography, influenced by factors such as the type of pornography viewed; the viewer's age, gender, and sexuality; preheld beliefs; and access to alternative sources of information [10,16,21,22]. How these factors influence the impact of pornography on knowledge, attitudes, and behaviors is beginning to be explored among subgroups of young people to understand risk factors for harm. A 2018 cross-sectional survey of 15- to 29-year-old heterosexual Australians (n=517) by our group found that young women (n=320) were significantly more likely than young men (n=197) to report having seen violent or degrading behaviors toward women [23]. We hypothesized that gendered experiences may influence how young audiences interpret pornographic images rather than indicate pornography preferences [23]. A 2015 study of economically disadvantaged, urban-residing black and Hispanic youth (n=72) identified an association between regular pornography use and dating violence [10]. Other studies have found that pornography can play an important role in the sexual development of lesbian, gay, bisexual, transsexual, queer, and intersex plus (LGBTQI+) young people in the context of limited access to alternative representations and overrepresentation of heteronormative sex [24,25]. However, few studies have explored how this may negatively influence ideas about sex, gender, and sexuality for young LGBTQI+ people exposed to fetishized representations of sexuality [26]. In summary, there may be young people who are at greater risk of harms from pornography as a result of the sociocultural and environmental contexts in which they view it. For instance, young people who do not consistently attend school, experience family conflict, identify as gender or sexuality diverse, or are from culturally and linguistically diverse backgrounds may have specific experiences and less access to relevant sexuality and relationships education, including pornography literacy. Understanding how these subgroups of young people experience pornography and its effects is vital to creating effective health promotion responses.

Pornography Literacy

Despite the evidence of the impacts of pornography on young people being underdeveloped, interventions to reduce its potential harms are being funded and produced [27-30]. Pornography literacy has been advocated as a harm reduction approach. On the basis of the media literacy theory [31], the key aims of pornography literacy approaches are to teach young people skills to critically analyze the messages in pornography, to increase their understanding of risks of exposure, and to encourage them to hold critical attitudes toward viewing the content and its messages [21,29,31,32]. However, there is little consensus about what pornography literacy education programs should include or how they should be delivered [33]. One of the few studies that explored the pornography literacy needs of young people used participatory research methods to identify information needs among Irish young adults aged 18 to 29 years [33]. Dawson identified pornography literacy priorities, including communication and consent, body image, reality of real-life sex, pleasure, information about the fetishization of gender and sexually diverse groups, and reducing shame over pornography use [33].

Importantly, current literature is inconclusive about the effectiveness of pornography literacy to reduce pornography-related harms among young people [32,34,35]. A small body of the literature suggests that participation in media literacy may help young people to develop skills to critique media messages and develop realistic expectations [33]. For instance, a representative panel survey of Dutch adolescents found an association between receiving school-based pornography literacy education and reduced strength in the relationship between pornography and stereotypical sexual attitudes [35]. However, the study did not define pornography literacy, relying on self-reporting of receiving any school-based education about pornography [35]. A recent preliminary study of 24 college students (aged 15-24 years) assessed the feasibility and effectiveness of a class-based pornography literacy program; it found that pornography-related knowledge increased posttest and some attitudinal shifts resulted [32]. These programs have been implemented and evaluated among young people engaged through formal in-person education and, thus, may not be applicable to those not in formal education settings. To date, there are no publicly available online or digital pornography literacy resources that specifically target vulnerable young people. Research has suggested that online resources could complement and enhance information engagement [33] and are most valuable to young people who lack alternative information, including those with diverse gender and sexual identities [36]. As such, understanding the experience, needs, and wants of subgroups that may be particularly vulnerable to the effects of viewing pornography will assist in developing responsive and equitable digital resources.

Human-Centered Design

Human-centered design and design thinking approaches involve prioritizing users throughout the creation of a product or service rather than at the start or end [37,38]. As a tool of human-centered design and design thinking, co-design methods involve bringing together key users, designers, and subject matter experts who participate in iterative workshops to understand and empathize with users. This enables them to define key issues and opportunities, to develop ideas to overcome issues, and to design solutions. Using creative and generative activities enables engagement with lived experiences and perspectives in ways that can overcome limitations of verbal communication where topics are complex or sensitive [39,40]. If implemented well, co-design methods can provide participants with time and space to actually think about the issue as they create, form, and question opinions and reflect on theirs and others' experiences in a safe and productive setting. Creating, role playing, and interaction between facilitators and participants can shift power dynamics and create trust and mutuality to understand and define sensitive issues. Importantly, involving users who are experts of their own experience as co-designers of potential solutions can help overcome the limitations of relying on traditional formative research methods to create health interventions [41].

Despite the seemingly obvious alignment of design and health disciplines, recognized challenges for cross-disciplinary collaboration exist [40,42]. For instance, it has been acknowledged that health research principles such as control, generalizability, reduction of bias, and replicability are in contrast with co-design processes that are necessarily creative, iterative, and flexible to account for the needs of users [40]. Furthermore, challenges of bringing together cross-disciplinary expertise central to co-design processes, such as time, resources, and willingness to compromise on disciplinary approaches could influence those considering this approach. Even the process of communicating co-design processes and findings falls outside of qualitative research conventions. Despite these challenges, the potential for the co-design method to radically contribute to the field of public health and health promotion is immense. For instance, there has been growing recognition of the potential for this method to increase the effectiveness of health promotion practices, particularly where issues are complex, sensitive, and ill-defined [37,39,42,43]. This being the case, these methods are particularly relevant for addressing the impact of web-based pornography on vulnerable young people's sexual health and their preferences for pornography literacy education.

This Study

This study aimed to co-design a relevant, usable, and acceptable digital prototype to improve pornography literacy among vulnerable young people. The project involved a cross-disciplinary co-design, including vulnerable young people, service providers, public health researchers, design researchers, user interface designers, and web developers.

Methods

Participants

A total of 17 young people aged between 14 and 23 years participated in the multiphase co-design workshops. The participants included 5 young women, 1 gender nonbinary person, and 11 young men.

Procedure

Young people were eligible if they were currently engaged with one of our partner youth service providers, were aged 14 to 24 years, and had experiences or identified with of at least one of the following: fragmented school attendance or disengagement from mainstream education (irregular attendance or participation in alternative curriculums such as vocational or applied learning); limited access to relevant education on sex, sexuality, and relationships (ie, for young people who identify as LGBTQI+); experience of family conflict or breakdown; and culturally and linguistically diverse (CALD) backgrounds.

We partnered with youth services, education providers, and primary health care providers to embed us in their existing programs for young people. This included youth groups and alternative education programs for young people disengaged from mainstream schools. A written information and consent form was developed and was accompanied by a short video to facilitate information and consent for people with low literacy skills. Project information was housed on a website to ensure that community members and potential co-designers could access further information about the project or contact our team. Informed consent was obtained from participants aged 16 years and older. For participants aged 15 years, we followed consenting practices for mature minors [44] of partner organizations, which involved independently consenting those deemed mature minors and seeking parental consent for others. Parental consent was obtained for participants younger than 15 years. The participants received an Aus $40 (US $29) gift voucher for every 3-hour workshop, with a bonus Aus $20 (US $12.11) voucher for those who attended all workshops. Facilitators, participants, and youth workers communicated about any sensitive or traumatic issues that may have emerged during activities. Participants were given the opportunity to change activities, observe, or leave it any time and to debrief or access relevant services after each workshop. Ethics approval for the research was granted by the Alfred Health Human Research Ethics Committee (373/18).

Participant Characteristics

A total of 17 young people participated in the co-design workshops, including 5 females, 1 gender nonbinary person, and 11 males. Participants ranged in age from 14 to 23 years (mean 16.5 years). Furthermore, 6 participants were from CALD backgrounds (English as a second language), 5 identified as LGBTQI+, and 2 were living with an intellectual disability. Over one-third of our sample was disengaged from mainstream schooling and had experiences of fragmented education. A similar proportion had experiences of family conflict or breakdown.

Co-Design Workshops

We conducted a total of 12 co-design workshops across 3 groups. Workshops were designed across 4 phases adapted from design thinking [37]: *understand*, *define*, *ideate*, and *design*. Workshops were conducted in known and safe spaces, including schools, community centers, and youth services. Two design researchers (1 male and 1 female) and a public health researcher (female) facilitated all the workshops. Youth workers and teachers were present on site but not in the room during the workshops. Each workshop ran for 3 hours. Workshops were audio-recorded, and note-takers recorded key conversations, observations, and interactions.

The first workshop focused on introducing the process, building trust and rapport, creating shared boundaries and safety, and beginning the process of getting to know our users and their information preferences. Activities were designed to create opportunities even during the introduction and icebreakers to understand more about our user's lives, aspirations, motivations, and challenges to help us walk in their shoes. The second workshop focused on creating opportunities for our participants to reflect on their thoughts, feelings, and experiences of sex, relationships, and pornography and to enable them to define key issues or problems, hence identifying information needs. We were conscious of creating space for participants to speak openly, using their own language about issues they may feel unsafe or embarrassed to talk about in the workshop setting. We designed activities that could be fully engaged with either anonymously or in the third person. The aim of the third workshop was to review young participants' definition of the problem and position our participants as *solution designers*. This involved building participants' understanding of design processes and tools and belief in themselves as *designers*, while simultaneously exploring their preferences for solutions. Activities facilitated the ideation of *blue-sky* ideas to help address issues defined throughout the previous workshops. The aim of the fourth workshop was to enable participants to develop their ideas into digital design solutions and to understand their brand values and aesthetic preferences for the user interface design. Participants used digital wireframes to map out their solution, features, and interactivity.

Workshop Activities

During activity design, we considered our participants' interests, abilities, and challenges and, our design practices. Activities prioritized creativity, shared understanding, empathy, and safety. In the first workshop, participants created personas of other young people that could be used for third-person discussions to avoid over disclosure. Across the workshops, each activity was built on the topics and outcomes from previous activities. This allowed participants to work into this complex issue and to increase trust and engagement as the activities progressed. Activities were generative to create spaces for participants to embrace creativity and to reflect on the issues [45]. Informed by previous co-design projects engaging with young people [39,43], we used a range of creative tools, including Lego, collage (Figure 1), community-mapping, drawing, music, and movie making. For instance, one activity involved participants creating a storyboard based on a scenario related to sex, relationships, or porn. They used personas to explore what the people in the scenario thought, felt, and did. Participants used Lego to create each scene of their movie along with speech bubbles. Each small group used a preloaded *stop motion* video app [46] to create a short movie, which they recorded and played back to the group for discussion. Activities positioned participants as designers throughout the process from understanding and defining the issue right through to the design of actual solutions. We conducted solution prototyping using pop culture references, such as superheroes, to leverage universal recognition and engagement with the Marvel comic franchise. A full description of our activity development and implementation is described in a research protocol and can be accessed on request.

Figure 1. Hot or Not? Collage.

Data Analysis

Data generated from activities included (1) the products of these activities that were considered *assets* that we could analyze; (2) our debrief notes that recorded conversations, interactions, and observations; and (3) verbal recordings of participants' discussions during the activities. All findings are developed based on insights drawn from across activities.

Co-Design Synthesis

An iterative human-centered design data synthesis using *affinity mapping* guided the analysis of data [47]. Step 1 involved extracting data from the assets the participants created, workshop recordings, transcription notes, and debrief notes. Step 2 involved sorting the data into several deductive themes that were clustered and coded together to establish subthemes. This step involved building consensus among the team through disagreement and debate. Step 3 involved *collaborative sense-making* where the research team interpreted the themes and subthemes into design insights [47]. For instance, the theme of youth trauma was evident across the workshop; in our data synthesis phase, we translated this into a design recommendation of needing our resource to be trauma-informed, which we placed under the *information preferences* branch of design considerations. We identified how our insights were connected to create a picture of our users' needs and challenges for the design. Each stage of the synthesis process included researchers AD and SM, and the final steps also included researcher CW and members of the design team.

Prototype Design

Human-centered design practices guided our process for bringing together the needs and wants of our users with best practice in accessible design and health promotion. The multidisciplinary team of designers and researchers were involved in the final stages of the synthesis and then participated in a process called Design Sprints. The process included ideating solutions based on user personas (available on request), on participants' design solutions created during co-design workshops, and on the response to the information needs and preferences of each user persona. Multiple rounds of ideation and design took place before a set of design recommendations for the prototype was developed. On the basis of the feedback from participants, key stakeholders, and external subject matter experts, these recommendations were developed into a functioning digital prototype of the solution for testing and evaluation (to take place in an independent study).

Results

Design Considerations

A summary of key design considerations (*information needs* and *information preferences)* identified through data synthesis and translation of co-design insights have been outlined as follows, after which the full description of the final prototype design has been provided.

Information Needs

Pornography Literacy

Participants defined pornography as *anything you can see on a porn site*, including porn videos, images, explicit pop-up advertisements (eg, for penis enlargements), and third-party services (eg, for paid sex). Activities and discussion among our users illustrated that they understood pornography as obviously unrealistic and could already critique its messages. Young women identified concerns over the representation of women's bodies, gender stereotypes such as *women who like sex are sluts*, and the reinforcement of rape myth acceptance. In contrast, young men and young transmen commonly identified problematic messages about bodies, including male bodies, male sexual performance, and normalization of the objectification of women, such as their representation as *i* (Figure 2). Despite being able to identify problematic messages, young men did not link them to their own attitudes or behaviors. Rather, some explicitly stated that knowing pornography was "fake" made them "porn proof." Others explained "we don't all watch Harry Potter and want to become wizards" to articulate to the group how their understanding of pornography as *fantasy* meant that they could watch it without having their attitudes or behaviors affected.

In contrast to these claims, young men created scenarios in which watching pornography did impact their behavior. For instance, a pair of participants created a stop-motion Lego movie illustrating how a character watched pornography and, after viewing anal sex, decides to try it with his girlfriend who is uncertain about it (Figure 3). In the scenario, the girl asks the guy to stop because she does not feel comfortable. He complies, only saying "I saw it online," and no harm is perceived to be done. The scenario and subsequent discussions revealed a contradiction to our participants' statements that watching porn did not affect their attitudes or behavior in any way and potential barriers to addressing impacts by focusing on the potential harms from porn.

Figure 2. Porn world vs. real world activity creation.

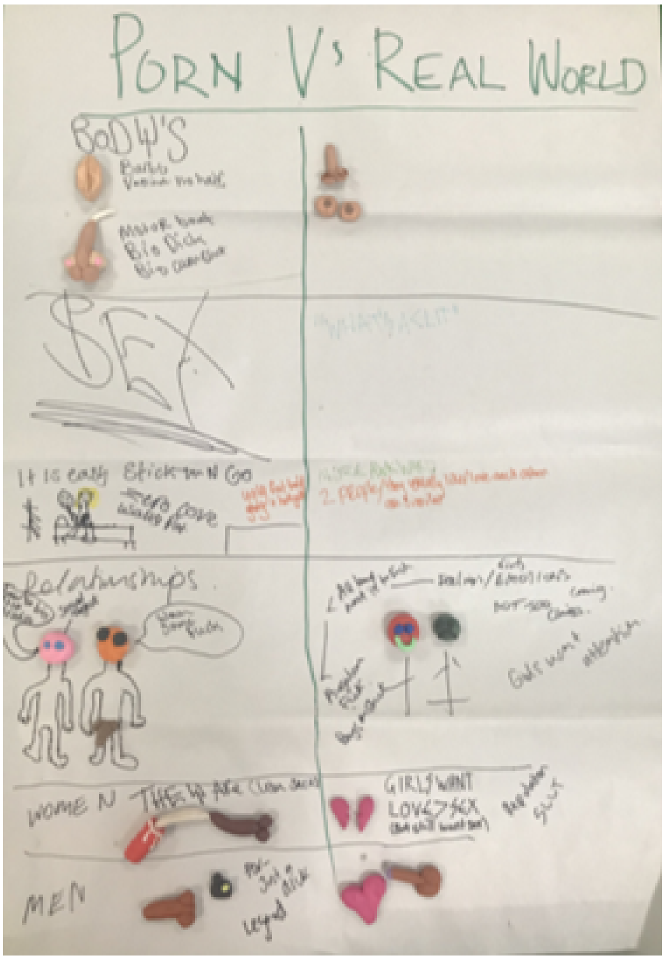
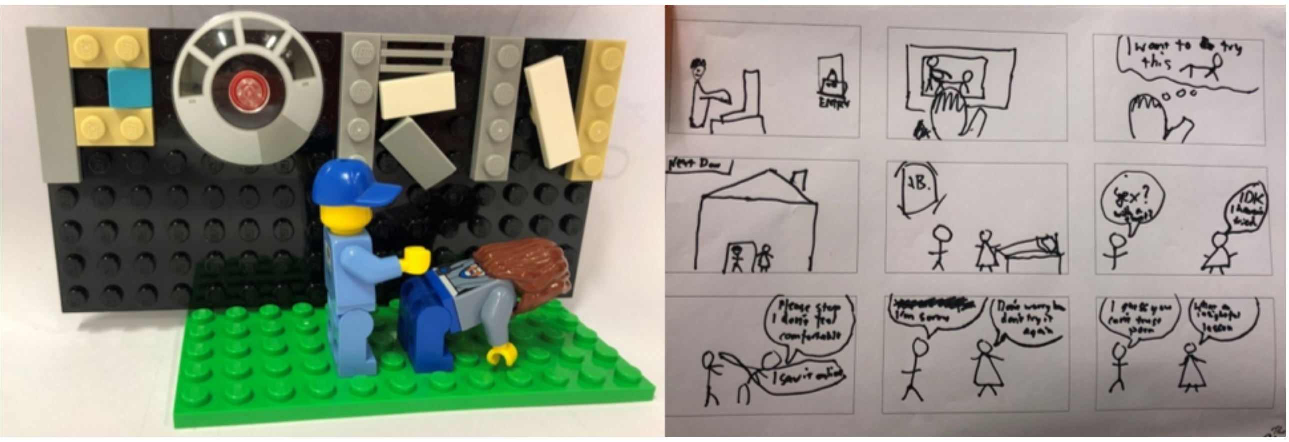


Figure 3. Stop motion movie scene and storyboard activity creation.



Beyond a Problem With Porn

Activities revealed that our users lacked information and skills to be able to contrast problematic messages from pornography with alternative ways of thinking or behaving. For instance, although participants could identify that representations of pleasure and consent in pornography were limited, there was little consensus among the groups about how to make sex pleasurable for all people and what constitutes consent or ethical sexual practices. This was particularly problematic for LGBTQI+ young people who experienced even less access to relevant representations of sex and sexuality outside of porn. Specific information gaps identified across groups included pleasurable sex positions, sexual pleasure for young women and trans young people, sexual pressure, consent, safe sex, and healthy relationships. Without alternative information and frameworks to guide their decision making and preferences, our participants were concerned with their ability to navigate both the "basics" and the "grey" areas of sex and relationships.

Furthermore, our users identified how messages from pornography can be compounded by messages they are exposed to in other domains. For young women in particular, problematic messages that they are exposed to in pornography are not always different from those in popular culture and society. For instance, some talked about the messages they are exposed to on Instagram about the ideal female body and sexuality. Some young men also articulated concerns about unhealthy behaviors they had seen modeled by family members such as a "drunk uncle" who makes people feel uncomfortable with his behavior. These insights highlight potential problems with creating a resource centered on pornography as an issue without addressing the information gaps and underlying social norms that permeate young people's lives.

Information Preferences

Reactive Information Seeking

In general, our users were unlikely to proactively seek out information, especially about sex, relationships, and porn, unless prompted by experience or contradiction to preheld knowledge. Rather, they wait until they have a negative or confusing experience or until their beliefs are contradicted to *Google it* This was illustrated during workshops when participants would often talk to each other about the issues raised through activities. On one occasion, 2 participants debated if girls can orgasm during penetrative sex. After discussion and disagreement, they pulled out their phones and searched for an answer. Both confirmed their own assumptions by finding something online to illustrate their point, highlighting that once they seek out information, many look to confirm or validate their existing opinions or beliefs. Importantly, neither participant found quality or evidence-based information; we found that few participants had the knowledge or skills to research high-quality information in practice. This has important implications for designing a resource on this topic.

Trauma-Informed

Participant's experiences of trauma affected their engagement with these sensitive topics. Some who had experienced sexually related trauma found overtly sex-positive sex education content retraumatizing. Other participants specified finding visual and descriptive content about sexually transmitted infections and detailed real-life stories of sexual violence to be traumatizing. As stated by a young man who said:

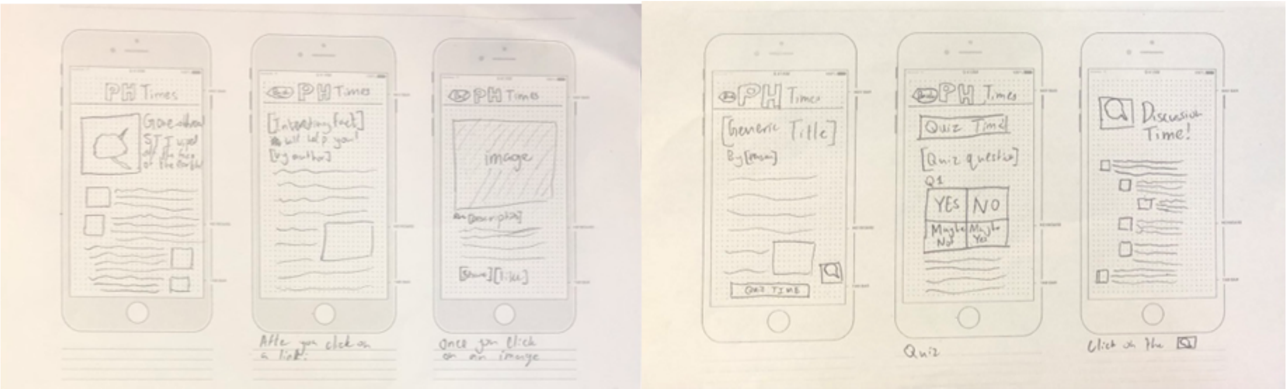
> *You don't want to see like diseased bodies (sic). It's just gross. But at the same time, you want to know about what happens. It's the same with like stories about sexual assault. I'm sorry but they are just hard to read sometimes.*

Although there was an understanding of the need to be aware of behaviors and consequences, young people's personal and online experiences of this educational content sometimes resulted in them withdrawing from engagement altogether. The sentiments demonstrated the fine balance between being direct and positive about sex without traumatizing or downplaying the sometimes negative experiences of sex among our users.

Prototype

On the basis of the results described earlier, a high-fidelity digital prototype was designed by a cross-disciplinary team of researchers and designers. A digital prototype is a way of creating a working example of a product that is tangible enough that it can be tested with a larger population and further developed based on new learnings. The design processes leveraged the solution designs (Figure 4), needs, and wants of co-design participants that were aligned with health promotion and design accessibility best practice [48]. The digital prototype developed, called *The Gist,* is described in the following sections with reference to the rationale for content (relevance and acceptability) and design (usability and accessibility).

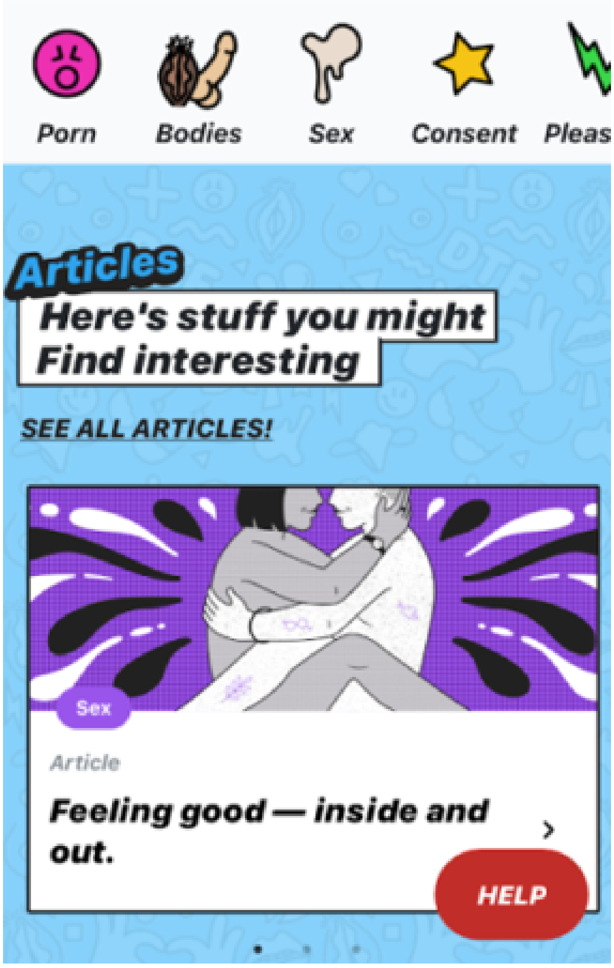
Figure 4. A digital wireframe of a solution concept created by a young co-designer during prototyping.



The Gist

The Gist (Figure 5) is an interactive mobile-first web-based app prototype. The aim of the design is to provide vulnerable young people with an interesting and easy way to take control of their sexual health and well-being in the context of normalized pornography use.

Figure 5. The Gist web-based mobile application prototype home screen.



Here's stuff you might Find interesting

Feeling good — inside and out.

Content

The Gist content provides alternative information that our users want and are not getting from porn, society, or mainstream education; it aims to develop healthy ideas about relationships, sex, and bodies. Drawing on concepts underpinning ethical sex, as described by Carmody and Ovendon [49], the resource aims to provide practical steps that users can follow to develop skills and preferences and help them to consider how their choices impact on themselves and others. For instance, practical information about sex is provided to help them develop an understanding of their own needs and their partners' needs. This content is linked with steps they can take to identify, understand,

and communicate these needs. Critical pornography literacy is built into messages by contrasting this alternative information with problematic messages present in pornography and society. This shifts the focus away from the *problems with porn* toward meeting our users' identified needs, while still enhancing the critical literacy of problematic representations of sex, sexuality, and gender. This includes information about sex positions, finding out what you like, sexual pleasure, sexual arousal, bodies, relationships, consent, and porn. The content balances direct, challenging, and informative articles with opportunities to build and test skills through interactive elements such as quizzes (Figure 6). Content is presented in gender- and sexuality-neutral ways in response to enable young people to explore information in an inclusive environment.

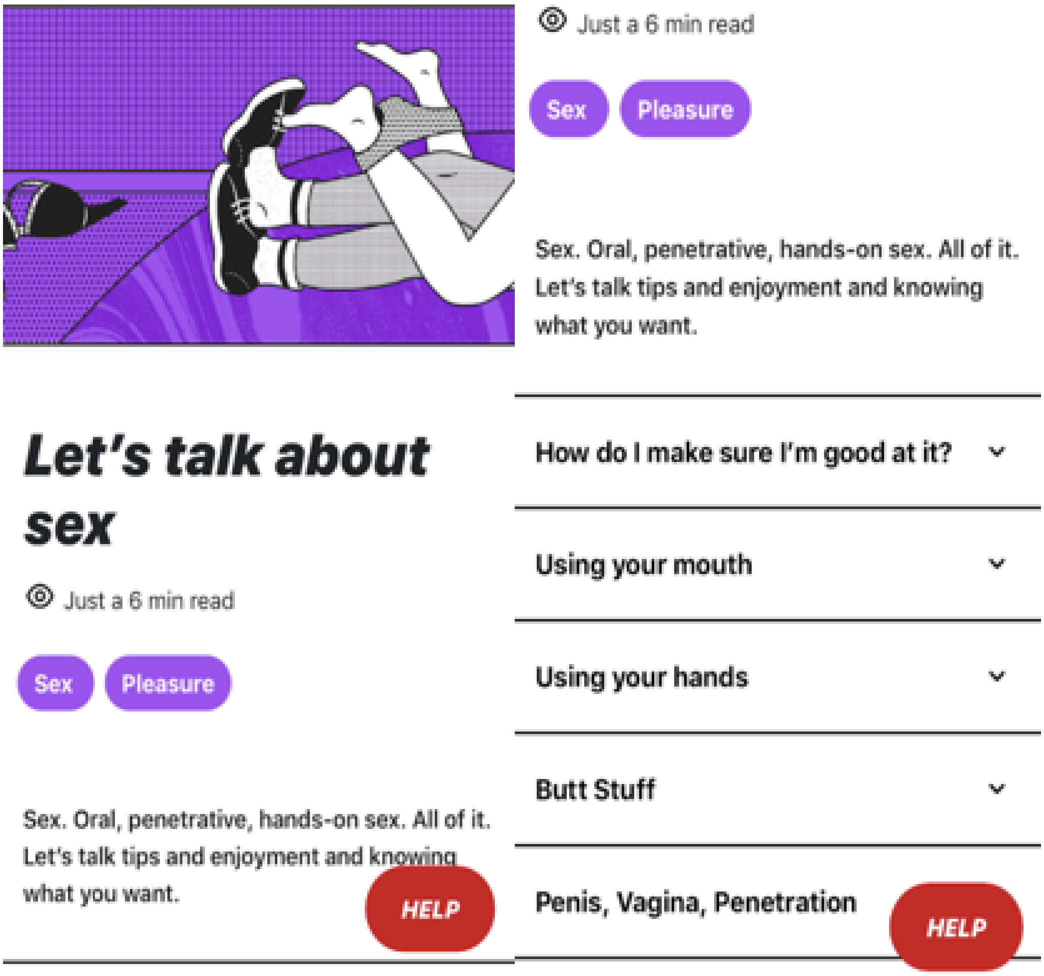
Figure 6. The Gist 'Debunked' activity to challenge pre-held knowledge and attitudes.

User Interface and Experience Design

The Gist brand is designed with a bold typography in keeping with our audiences' desire for alternative and nonmainstream products, and a *retro* 90s design feel. The name, *The Gist*, refers to the process of getting to know the idea or main point of a concept. It was selected for the prototype through testing with co-design participants. The user interface leverages our users' reactive information-seeking habits, and quick content for browsing aims to surprise or contradict our users' ideas and experiences motivating them to seek out more information. Short *gists* or facts are linked through to deeper exploration of the issues to enable in-app learning to progress once a user is engaged. Multiformat content also leverages user experience (UX) and interaction patterns from apps our users frequently engage with such as Spotify, Instagram Stories, and YouTube. This enhances the familiarity and usability of The Gist for our target population. Capitalizing on our users current UX patterns, the app enables continuous scrolling for breadth of engagement and continuous links to further content for depth of engagement [48]. Detailed articles utilize progressive disclosure (informative headlines with expandable information sections; for example, Figure 7). Users are provided with staggered content that can be engaged with if elected. This allows young people to take control of what they see and to elect not be exposed to information that they are uncomfortable with. This is based on insights from the co-design that suggest that lived experience and developmental stage are more important factors than age in relation to information needs, relevance, and acceptability.

The Gist web-based app does not include commenting features or social media integration other than sharing a quiz or an article directly with someone. This reduces the capacity for young people to interact negatively with each other and reflects our co-design research and previous studies suggesting that young people are unlikely to actively share any content relating to sexual health on social media [41]. Although the prototype design can be developed as a standalone, direct-to-consumer digital resource; we recommended that implementation be embedded within a multipronged health promotion approach such as service-delivered workshops. This is in response to the positive reaction that our co-design workshops were met with and the participants explicitly stating that they valued the opportunity to direct conversations with adults on sex and relationships. Scaffolding the digital product with service-delivered face-to-face workshops would also act as an entry point to the digital solution to enhance reach and engagement, making a digital solution more feasible.

Figure 7. The Gist article structure utilizing progressive disclosure to increase user control.



Let's talk about sex

Sex. Oral, penetrative, hands-on sex. All of it. Let's talk tips and enjoyment and knowing what you want.

Sex. Oral, penetrative, hands-on sex. All of it. Let's talk tips and enjoyment and knowing what you want.

Discussion

Principal Findings

The findings of our co-design workshops identify unmet sexual health and pornography literacy information needs among our users, their information-seeking patterns, and preferences for a digital resource. Through this process, we were able to develop The Gist prototype with young co-designers. The detailed and revealing nature of our findings suggests that using participatory research methods such as co-design can overcome sociocultural and structural barriers to engaging vulnerable young people in research on sensitive health issues. Some notable findings from this cross-disciplinary approach are discussed in the following sections in the context of current literature and directions for future research.

Unmet Information Needs and the Limitations of Current Pornography Literacy Approaches

Most formal pornography literacy programs focus on the critical awareness of pornographic content and the industry [32]. We found that our users understood pornography as obviously unrealistic and could already critique its messages. Consistent with findings from the study by Dawson et al [33], a key message of this study is that effective pornography literacy programs must expand their focus on problematic messages from pornography and go beyond critical media literacy to include alternative information to help young audiences to navigate sexuality and relationships in their lives.

Few resources exist that directly contrast information about sex with messages from pornography and society [50-52]. There are some notable exceptions that include a cross-section of relevant topics [53]. However, the usability and acceptability of the resource among subgroups of young people are unknown. Our findings suggest that there is potential to embed pornography literacy within practical information about ethical sex and relationships to help young people transfer critical literacy into alternate choices and preferences in real life, but that this approach must be evaluated [49]. Any approach must balance sex-positive language and framing with acknowledgment of the negative lived experiences of some young people who require sensitively delivered information acknowledging these experiences. Although we included safeguarding approaches such as progressive disclosure and trauma-informed language, further research would strengthen the approach to pornography literacy programs.

A large body of the literature has identified the information deficit that young people experience as a result of current education approaches that fail to adequately teach about pleasure, sexual discovery, or self-representation [21,54,55]. We found that for vulnerable young people, there were critical unmet information needs in relation to the basics of navigating sex and relationships that compounded fetishized representations in pornography [33]. Thus, having relevant and accessible alternative information may be even more vital for vulnerable young people whose basic needs may not be met by schools, parents, or peers [25]. Importantly, our findings reinforce the need for longitudinal research to understand the specific impacts of pornography on the lives of subgroups of young people, including disengaged LGBTQI+ and CALD groups, if their needs are to be met.

Co-Design and Digital Health Promotion

Research has illustrated limitations to traditional approaches to intervention design, which involve engagement with target users in research to inform content development by subject matter or technical experts rather than by engaging end users throughout the process as equal partners in design [39,56,57]. Where topics are particularly complex or sensitive, asking young people to engage in one-on-one interviews or even focus groups may be ethically sensitive and/or increase social desirability bias. Using creative and generative activities, we observed an extremely high level of engagement among young people, even on difficult topics that sometimes included traumatic experiences. Creating a safe space where participants could use creation to express an experience, rather than having to speak directly about it, enabled them to have their perspectives included and to better understand the underlying issues and design a more inclusive solution. The study provides an example of how co-design methods are responsive to sensitive and complex health issues and are particularly suitable for designing with rather than for young people [40,43]. More work needs to be done evaluating the effectiveness of co-designed interventions compared with evidence-based expert-designed approaches.

Limitations

These findings should be considered within the context of study limitations. Co-design with *extreme* or *hard-to-reach* users generates products that can appeal to broader populations [39,43,58]; however, co-design research needs to be undertaken to test and adapt the prototype for specific audiences. We could not report quotes and attributions in this publication because of ethical concerns that service providers who helped with recruitment may be able to identify participants. Despite these limitations, the study provides important findings relating to pornography literacy and intervention design because of high-quality, in-depth, and reoccurring engagement with this vulnerable population.

Conclusions

The study findings include important recommendations for the conceptualization of online pornography literacy programs and opportunities for cross-disciplinary research to address complex and emerging health issues. If pornography literacy programs focus on pornography without providing young people with alternative information about sex and relationships, they may not be effective in reducing harm. This study highlights the potential for pornography literacy to be embedded within real and practical information about having sex, navigating sexuality, and healthy relationships. The cross-disciplinary approach reported in this study demonstrates that it is possible to collaboratively co-design a pornography literacy resource combining the needs and wants of vulnerable young people with best practice in design and health promotion.

Acknowledgments

The project was funded by the Victorian Health Promotion Foundation, Melbourne, Australia. Prototype content was developed with technical expertise from Jenny Walsh. Illustration and production were delivered in partnership with Cat Wall and Bonnie Eichelberger. AD was supported by the National Health and Medical Research Council postdoctoral scholarship. CW, ML, PD, and MH were supported by the National Health and Medical Research Council fellowships. The project was delivered with collaboration between the Burnet Institute and Sheda and in partnership with Stonnington Youth Services, Hester Hornbrook Academy, and Star Health. The authors gratefully acknowledge the contribution to this work of the Victorian Operational Infrastructure Support Program received by the Burnet Institute.

Conflicts of Interest

None declared.

References

1. Owens EW, Behun RJ, Manning JC, Reid RC. The impact of internet pornography on adolescents: a review of the research. Sex Addict Compulsivity 2012 Jan;19(1-2):99-122. [doi: 10.1080/10720162.2012.660431]

2. Mattebo M, Larsson M, Tydén T, Häggström-Nordin E. Professionals' perceptions of the effect of pornography on Swedish adolescents. Public Health Nurs 2014;31(3):196-205. [doi: 10.1111/phn.12058] [Medline: 24720654]
3. Gorman S, Monk-Turner E, Fish JN. Free adult internet web sites: how prevalent are degrading acts? Gend Issues 2010 Nov 10;27(3-4):131-145. [doi: 10.1007/s12147-010-9095-7]
4. Klaassen MJ, Peter J. Gender (In)equality in internet pornography: a content analysis of popular pornographic internet videos. J Sex Res 2015;52(7):721-735. [doi: 10.1080/00224499.2014.976781] [Medline: 25420868]
5. Vannier SA, Currie AB, O'Sullivan LF. Schoolgirls and soccer moms: a content analysis of free 'teen' and 'MILF' online pornography. J Sex Res 2014;51(3):253-264. [doi: 10.1080/00224499.2013.829795] [Medline: 24228745]
6. Bridges AJ, Wosnitzer R, Scharrer E, Sun C, Liberman R. Aggression and sexual behavior in best-selling pornography videos: a content analysis update. Violence Against Women 2010 Oct;16(10):1065-1085. [doi: 10.1177/1077801210382866] [Medline: 20980228]
7. Lim MS, Agius PA, Carrotte ER, Vella AM, Hellard ME. Young Australians' use of pornography and associations with sexual risk behaviours. Aust N Z J Public Health 2017 Aug;41(4):438-443. [doi: 10.1111/1753-6405.12678] [Medline: 28664609]
8. Harkness EL, Mullan BM, Mullan BM, Blaszczynski A. Association between pornography use and sexual risk behaviors in adult consumers: a systematic review. Cyberpsychol Behav Soc Netw 2015 Feb;18(2):59-71. [doi: 10.1089/cyber.2014.0343] [Medline: 25587721]
9. Luder M, Pittet I, Berchtold A, Akré C, Michaud P, Surís JC. Associations between online pornography and sexual behavior among adolescents: myth or reality? Arch Sex Behav 2011 Oct;40(5):1027-1035. [doi: 10.1007/s10508-010-9714-0] [Medline: 21290259]
10. Rothman EF, Adhia A. Adolescent pornography use and dating violence among a sample of primarily black and Hispanic, urban-residing, underage youth. Behav Sci (Basel) 2015 Dec 23;6(1) [FREE Full text] [doi: 10.3390/bs6010001] [Medline: 26703744]
11. Jochen P, Valkenburg PM. Adolescents' exposure to sexually explicit internet material and notions of women as sex objects: assessing causality and underlying processes. J Commun 2009 Sep;59(3):407-433. [doi: 10.1111/j.1460-2466.2009.01422.x]
12. Ybarra ML, Strasburger VC, Mitchell KJ. Sexual media exposure, sexual behavior, and sexual violence victimization in adolescence. Clin Pediatr (Phila) 2014 Nov;53(13):1239-1247. [doi: 10.1177/0009922814538700] [Medline: 24928575]
13. Braun-Courville DK, Rojas M. Exposure to sexually explicit web sites and adolescent sexual attitudes and behaviors. J Adolesc Health 2009 Aug;45(2):156-162. [doi: 10.1016/j.jadohealth.2008.12.004] [Medline: 19628142]
14. Häggström-Nordin E, Hanson U, Tydén T. Associations between pornography consumption and sexual practices among adolescents in Sweden. Int J STD AIDS 2005 Feb;16(2):102-107. [doi: 10.1258/0956462053057512] [Medline: 15807936]
15. Brown JD, L'Engle KL. X-rated: sexual attitudes and behaviors associated with US early adolescents' exposure to sexually explicit media. Commun Res 2009 Feb;36(1):129-151. [doi: 10.1177/0093650208326465]
16. Mattebo M, Tydén T, Häggström-Nordin E, Nilsson KW, Larsson M. Pornography consumption, sexual experiences, lifestyles, and self-rated health among male adolescents in Sweden. J Dev Behav Pediatr 2013 Sep;34(7):460-468. [doi: 10.1097/DBP.0b013e31829c44a2] [Medline: 23899659]
17. Lim MS, Carrotte ER, Hellard ME. The impact of pornography on gender-based violence, sexual health and well-being: what do we know? J Epidemiol Community Health 2016 Jan;70(1):3-5. [doi: 10.1136/jech-2015-205453] [Medline: 26022057]
18. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001 Aug;3(3):265-299. [doi: 10.1207/s1532785xmep0303_03]
19. Wright PJ. Mass media effects on youth sexual behavior assessing the claim for causality. Ann Int Commun Assoc 2011;35(1):343-385. [doi: 10.1080/23808985.2011.11679121]
20. Fritz N, Paul B. From orgasms to spanking: a content analysis of the agentic and objectifying sexual scripts in feminist, for women, and mainstream pornography. Sex Roles 2017;77(9-10):639-652. [doi: 10.1007/s11199-017-0759-6]
21. Albury K. Porn and sex education, porn as sex education. Porn Stud 2014 Mar 21;1(1-2):172-181. [doi: 10.1080/23268743.2013.863654]
22. Mckee A. Does pornography harm young people? Aust J Commun 2010;37(1):17-36 [FREE Full text]
23. Davis AC, Carrotte ER, Hellard ME, Lim MSC. What behaviors do young heterosexual Australians see in pornography? A cross-sectional study. J Sex Res 2018;55(3):310-319. [doi: 10.1080/00224499.2017.1417350] [Medline: 29333877]
24. Kubicek K, Beyer WJ, Weiss G, Iverson E, Kipke MD. In the dark: young men's stories of sexual initiation in the absence of relevant sexual health information. Health Educ Behav 2010 Apr;37(2):243-263 [FREE Full text] [doi: 10.1177/1090198109339993] [Medline: 19574587]
25. Davis A, Carrotte E, Hellard M, Temple-Smith M, Lim M. Pornography as a source of education about sex and sexuality among a sample of 15–29 year old Australians. J Sex Med 2017 May;14(5):e272. [doi: 10.1016/j.jsxm.2017.04.685]
26. Dawson K. Educating Ireland: promoting porn literacy among parents and children. Porn Studies 2018 Nov 5;6(2):268-271. [doi: 10.1080/23268743.2018.1518727]
27. It's Time We Talked. What Can Schools Do to Address Pornography’s Influence? URL: http://www.itstimewetalked.com.au/schools/what-can-schools-do/ [accessed 2019-03-14]

28. It's Time We Talked. What Can You Do? URL: http://www.itstimewetalked.com.au/parents/what-can-you-do/ [accessed 2019-03-14]
29. It's Time We Talked. 2016. URL: http://www.itstimewetalked.com.au [accessed 2019-03-14]
30. The Australian Research Centre in Sex, Health and Society, La Trobe University. The Practical Guide to Love, Sex and Relationships. Melbourne, Australi; 2015. URL: http://www.lovesexrelationships.edu.au/ [accessed 2016-01-05]
31. Pinkleton BE, Austin EW, Chen YY, Cohen M. The role of media literacy in shaping adolescents' understanding of and responses to sexual portrayals in mass media. J Health Commun 2012;17(4):460-476. [doi: 10.1080/10810730.2011.635770] [Medline: 22273591]
32. Rothman EF, Adhia A, Christensen TT, Paruk J, Alder J, Daley N. A pornography literacy class for youth: results of a feasibility and efficacy pilot study. Am J Sex Educ 2018 Feb 22;13(1):1-17. [doi: 10.1080/15546128.2018.1437100]
33. Dawson K, Gabhainn SN, MacNeela P. Toward a model of porn literacy: core concepts, rationales, and approaches. J Sex Res 2020 Jan;57(1):1-15. [doi: 10.1080/00224499.2018.1556238] [Medline: 30624090]
34. Austin EW, Pinkleton BE, Chen YC, Austin BW. Processing of sexual media messages improves due to media literacy effects on perceived message desirability. Mass Commun Soc 2015 Feb 9;18(4):399-421. [doi: 10.1080/15205436.2014.1001909]
35. Vandenbosch L, van Oosten JM. The relationship between online pornography and the sexual objectification of women: the attenuating role of porn literacy education. J Commun 2017 Nov 1;67(6):1015-1036. [doi: 10.1111/jcom.12341]
36. Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb;29(1):147-157 [FREE Full text] [doi: 10.1093/her/cyt071] [Medline: 23861481]
37. Chan K. A design thinking mindset beyond the public health model. World Med Health Policy 2018 Mar 12;10(1):111-119. [doi: 10.1002/wmh3.253]
38. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
39. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth interventions with children and young people. Front Psych 2018;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]
40. Bazzano AN, Martin J, Hicks E, Faughnan M, Murphy L. Human-centred design in global health: A scoping review of applications and contexts. PLoS One 2017;12(11):e0186744 [FREE Full text] [doi: 10.1371/journal.pone.0186744] [Medline: 29091935]
41. Wright CJ, Davis A, Vella A, McGrath-Lester R, Lim MS. Young people’s needs and preferences for health resources focused on pornography and sharing of sexually explicit imagery. Pub Health Res Pract 2019:- in press. [doi: 10.17061/phrp29341912]
42. Chamberlain P. A disciplined approach to interdisciplinary working. Des Health 2019;2(2):191-193. [doi: 10.1080/24735132.2018.1542053]
43. Nakarada-Kordic I, Hayes N, Reay SD, Corbet C, Chan A. Co-designing for mental health: creative methods to engage young people experiencing psychosis. Des Health 2017;1(2):229-244. [doi: 10.1080/24735132.2017.1386954]
44. Rallis Legal. Research Governance Service Australia. 2017 Feb. Laws Relating to the Giving of Consent for Persons With Impaired Capacity to Provide Informed Consent to Participate in Research in Each Australian State and Territory URL: https://rgs.health.wa.gov.au/Documents/NHMRC_final_report_consent_and_guardianship_laws.pdf [accessed 2020-03-27]
45. Dahl DW, Chattopadhyay A, Gorn GJ. The importance of visualisation in concept design. Des Stud 2001 Jan;22(1):5-26 [FREE Full text] [doi: 10.1016/S0142-694X(99)00028-9]
46. Cateater L. App Store - Apple. 2019. Stop Motion Studio URL: https://itunes.apple.com/au/app/stop-motion-studio/id441651297?mt=8 [accessed 2020-06-30]
47. Pernice K. Neilson Norman Group. Online article Online: Neilson Norman Group; 2018 Feb 18. Affinity Diagramming for Collaboratively Sorting UX Findings and Design Ideas URL: https://www.nngroup.com/articles/affinity-diagram/ [accessed 2018-04-17]
48. Joyce A, Nielsen J. Nielson Norman Group. Online article Online: Nielson Norman Group; 2019 Mar 17. Teenager’s UX: Designing for Teens URL: https://www.nngroup.com/articles/usability-of-websites-for-teenagers/?lm=teenagers-on-the-web&pt=report [accessed 2018-08-21]
49. Carmody M, Ovenden G. Putting ethical sex into practice: sexual negotiation, gender and citizenship in the lives of young women and men. J Youth Stud 2013 Sep;16(6):792-807. [doi: 10.1080/13676261.2013.763916]
50. McCarthy O, Carswell K, Murray E, Free C, Stevenson F, Bailey JV. What young people want from a sexual health website: design and development of Sexunzipped. J Med Internet Res 2012 Oct 12;14(5):e127 [FREE Full text] [doi: 10.2196/jmir.2116] [Medline: 23060424]
51. Allen L. 'Looking at the Real Thing': young men, pornography, and sexuality education. Discourse Stud Cult Politics Educ 2006 Mar;27(1):69-83. [doi: 10.1080/01596300500510302]

52. Sundaram V, Sauntson H. Discursive silences: using critical linguistic and qualitative analysis to explore the continued absence of pleasure in sex and relationships education in England. Sex Educ 2015 Sep 4;16(3):240-254. [doi: 10.1080/14681811.2015.1058771]
53. Bish UK. URL: https://www.bishuk.com/ [accessed 2020-03-27]
54. Connell E. Desire as interruption: young women and sexuality education in Ontario, Canada. Sex Educ 2005 Aug;5(3):253-268. [doi: 10.1080/14681810500171433]
55. Fine M, McClelland S. Sexuality education and desire: still missing after all these years. Harv Educ Rev 2006;76(3):297-338. [doi: 10.17763/haer.76.3.w5042g23122n6703]
56. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
57. Davis AC, Wright CJC, Temple-Smith MJ, Hellard ME, Lim MS. A health education website developed to meet young people's information needs about web-based pornography and sharing of sexually explicit imagery (SCOPE): usability study. JMIR Form Res 2019 Aug 13;3(3):e12824 [FREE Full text] [doi: 10.2196/12824] [Medline: 31411140]
58. Dietrich T, Rundle-Thiele S, Schuster L, Connor J. Co-designing social marketing programs. J Soc Mark 2016 Jan 4;6(1):41-61. [doi: 10.1108/jsocm-01-2015-0004]

Abbreviations

CALD: culturally and linguistically diverse
LGBTQI+: lesbian, gay, bisexual, transsexual, queer, and intersex plus
UX: user experience

Edited by G Eysenbach; submitted 22.08.19; peer-reviewed by L Sun, H Thabrew, K Brown; comments to author 21.10.19; revised version received 17.11.19; accepted 24.02.20; published 01.06.20.

Please cite as:
Davis AC, Wright CJC, Murphy S, Dietze P, Temple-Smith MJ, Hellard ME, Lim MSC
A Digital Pornography Literacy Resource Co-Designed With Vulnerable Young People: Development of "The Gist"
J Med Internet Res 2020;22(6):e15964
URL: https://www.jmir.org/2020/6/e15964
doi:10.2196/15964
PMID:32348268

Original Paper

A Social Media–Based Support Group for Youth Living With HIV in Nigeria (SMART Connections): Randomized Controlled Trial

Lisa Dulli[1], MHS, PhD; Kathleen Ridgeway[1], MSPH; Catherine Packer[1], MSPH; Kate R Murray[1], MPH; Tolulope Mumuni[2], MSc; Kate F Plourde[1], MPH; Mario Chen[1], PhD; Adesola Olumide[2], MBBS, MPH, FWACP, FMCPH; Oladosu Ojengbede[2], MBBS, BSc, FMCOG, FWACS, FICS; Donna R McCarraher[1], MPH, PhD

[1]FHI 360, Durham, NC, United States

[2]Center for Population and Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Nigeria

Corresponding Author:
Lisa Dulli, MHS, PhD
FHI 360
359 Blackwell St, Suite 200
Durham, NC, 27701
United States
Phone: 1 9195447040 ext 11396
Email: ldulli@fhi360.org

Abstract

Background: Youth living with HIV (YLHIV) enrolled in HIV treatment experience higher loss to follow-up, suboptimal treatment adherence, and greater HIV-related mortality compared with younger children or adults. Despite poorer health outcomes, few interventions target youth specifically. Expanding access to mobile phone technology, in low- and middle-income countries (LMICs) in particular, has increased interest in using this technology to improve health outcomes. mHealth interventions may present innovative opportunities to improve adherence and retention among YLHIV in LMICs.

Objective: This study aimed to test the effectiveness of a structured support group intervention, Social Media to promote Adherence and Retention in Treatment (SMART) Connections, delivered through a social media platform, on HIV treatment retention among YLHIV aged 15 to 24 years and on secondary outcomes of antiretroviral therapy (ART) adherence, HIV knowledge, and social support.

Methods: We conducted a parallel, unblinded randomized controlled trial. YLHIV enrolled in HIV treatment for less than 12 months were randomized in a 1:1 ratio to receive SMART Connections (intervention) or standard of care alone (control). We collected data at baseline and endline through structured interviews and medical record extraction. We also conducted in-depth interviews with subsets of intervention group participants. The primary outcome was retention in HIV treatment. We conducted a time-to-event analysis examining time retained in treatment from study enrollment to the date the participant was no longer classified as active-on-treatment.

Results: A total of 349 YLHIV enrolled in the study and were randomly allocated to the intervention group (n=177) or control group (n=172). Our primary analysis included data from 324 participants at endline. The probability of being retained in treatment did not differ significantly between the 2 study arms during the study. Retention was high at endline, with 75.7% (112/163) of intervention group participants and 83.4% (126/161) of control group participants active on treatment. HIV-related knowledge was significantly better in the intervention group at endline, but no statistically significant differences were found for ART adherence or social support. Intervention group participants overwhelmingly reported that the intervention was useful, that they enjoyed taking part, and that they would recommend it to other YLHIV.

Conclusions: Our findings of improved HIV knowledge and high acceptability are encouraging, despite a lack of measurable effect on retention. Retention was greater than anticipated in both groups, likely a result of external efforts that began partway through the study. Qualitative data indicate that the SMART Connections intervention may have contributed to retention, adherence, and social support in ways that were not captured quantitatively. Web-based delivery of support group interventions can permit people to access information and other group members privately, when convenient, and without travel. Such digital health interventions may help fill critical gaps in services available for YLHIV.

Trial Registration: ClinicalTrials.gov NCT03516318; https://clinicaltrials.gov/ct2/show/NCT03516318

(J Med Internet Res 2020;22(6):e18343) doi:10.2196/18343

KEYWORDS

HIV; treatment adherence and compliance; youth; social support; mHealth; medication adherence; mobile phone

Introduction

Globally, young people aged 15 to 24 years account for more than 30% of new HIV infections, over 80% of whom live in sub-Saharan Africa (SSA); AIDS remains the leading cause of death among youth in SSA [1-6]. Despite a lack of age-disaggregated data on antiretroviral therapy (ART) coverage for youth, available data show that youth living with HIV (YLHIV) enrolled in HIV care experience higher loss to follow-up and suboptimal treatment adherence compared with younger children or adults [7-10]. YLHIV face many challenges to effective treatment, including stigma and fear of disclosure to others, lack of social support, lack of services designed to comprehensively meet their unique needs, and limited knowledge about the disease [7-10]. To achieve the Joint United Nations Programme on HIV and AIDS 95-95-95 goals by 2030, understanding and addressing the unique needs of YLHIV will be critical [11].

Although YLHIV experience poorer health outcomes, few interventions to improve these outcomes target youth specifically. Interventions to improve HIV-related outcomes implemented in low- and middle-income countries (LMICs) largely target adults and aim to improve ART adherence; fewer interventions target retention in care [12-14]. Support groups to improve health outcomes among people living with HIV (PLHIV), including YLHIV, are supported by some evidence to improve retention in care [15-22].

Expanding access to mobile phone technology has increased interest in using this technology to improve health outcomes. Digital health interventions such as mobile reminders and interactive voice or SMS responses have shown some effectiveness in improving adherence or retention among PLHIV in LMICs [23-31]. Although research has not examined these interventions specifically among YLHIV in LMICs, preliminary evidence in high-income countries suggests that such interventions are feasible and may impact ART adherence [32-35]. Furthermore, 2 recent studies in South Africa integrated social media into interventions for YLHIV to improve social support and found them to be acceptable and feasible [36,37]. Increasing internet use, feature phone, and smartphone availability in LMICs makes these interventions possible. Developing and testing interventions to improve YLHIV outcomes in SSA is urgently needed, and digital health interventions may present innovative opportunities to improve adherence and retention among youth in LMICs [12-14,38-40].

Nigeria is experiencing a generalized HIV epidemic and has the second largest global burden of the disease, with an estimated 3,438,442 people currently living with HIV [41]. As with other countries in the region, YLHIV in Nigeria experience high loss to follow-up from HIV treatment services, with the greatest losses occurring early in treatment [39,42]. Given increasing access to and use of mobile phone technology in Nigeria, digital health strategies have potential to help meet YLHIV informational and social support needs [43,44].

In response, we developed a structured support group intervention— Social Media to promote Adherence and Retention in Treatment (SMART) Connections—delivered through a social media platform. The intervention's aim is to improve HIV-related knowledge, social support, and ultimately, retention in HIV treatment and ART adherence among YLHIV. We conducted workshops with local stakeholders and youths to inform intervention design, followed by a feasibility and acceptability study with 41 adolescents aged 15 to 19 years in south-central Nigeria (results published elsewhere) [45]. In this study, we set out to test the effectiveness of the SMART Connections intervention on HIV treatment retention among youth aged 15 to 24 years. We also examined the effects of SMART Connections on secondary outcomes of ART adherence, HIV knowledge, and social support.

Methods

Design

We conducted a two-arm, parallel, randomized controlled trial. Participants were individually randomized in a 1:1 ratio to either the SMART Connections intervention (intervention group) or standard of care (control group). This was an open-label study with no blinding of the study staff or participants. We conducted face-to-face structured interviews with participants and extracted data from medical records at enrollment (baseline) and again at the completion of the intervention period (endline), approximately 6-9 months after enrollment. We also conducted in-depth interviews (IDIs) at endline with 2 subsets of intervention group participants: participants with high group participation and those with low group participation (classified by group facilitators). We obtained informed consent from youth aged 18 years and older; for youth aged 15 to 17 years, we obtained parental permission and adolescent assent at enrollment. This study was reviewed and approved by FHI 360's Protection of Human Subjects Committee, the University of Uyo Teaching Hospital's Institutional Health Research Ethical Committee, and the Cross River State Health Research Ethics Committee.

Intervention Description

SMART Connections was informed through workshops conducted with stakeholders and YLHIV in Akwa Ibom State, Nigeria, to gather input for design and content [12,13]. The intervention was designed to promote retention in HIV care by leveraging social support and improving HIV-related knowledge and treatment literacy [46,47]. The content of the structured support groups was adapted from an existing support group guide, Positive Connections, and delivered through secret Facebook groups [48]. The intervention was delivered over approximately 22 weeks (2 weeks per session) to groups of about 15 to 25 youths, with nearly daily activities (Figure 1).

The intervention guide is attached as a Multimedia Appendix 1.

Two community-based organizations were engaged to recruit support group facilitators and assist study staff in training them to deliver the intervention. Selected facilitators had received prior training to facilitate in-person support groups and were living with HIV themselves. Facilitators underwent a 5-day training and received an implementation guide, smartphone and monthly data allowance. Facilitators met monthly with the study staff to debrief on challenges they had encountered and provided support to one another.

Once a sufficient number of participants were recruited and randomized to form a support group, each group began with an in-person meeting, during which participants met one another and the facilitator and agreed-upon ground rules for participation. Facilitators also instructed participants on how to secure their phones (using a password or passcode), turn off Facebook notifications so that messages and posts would not show up on their phones and to log out of Facebook after each use to keep others from accessing their Facebook accounts.

Figure 1. Intervention design overview. YLHIV: youth living with HIV.

How it works

Safe Space: YLHIV ages 15 to 19 form secret Facebook groups.

Supportive Adult Role Model: Groups are led by a trained adult facilitator with lived HIV experience.

Social Activities: Including riddles, puzzles, and discussion about nonhealth topics to promote bonding.

Interactive Polls and Facilitated Discussions test knowledge and reinforce learning.

Word-of-the-Week and Key Message: Images provide definitions of core concepts.

11 Key Topics Covered: Disclosure, exploring feelings, nutrition and health, reproductive rights, positive health dignity and prevention, discrimination and rights, violence, communication and problem solving, and planning for your future.

Standard of Care Services

All study participants continued to receive standard services available to YLHIV in study facilities and communities. Standard health services included routine clinical care for HIV treatment, including viral load tests, active case management by community volunteers with intensive adherence support during the first 4 weeks of ART, and enhanced adherence counseling for patients with unsuppressed viral loads.

All study participants received a smartphone (valued at US $65) and a monthly data bundle (US $3.5 equivalent to 1 gigabyte of data) for the duration of the intervention.

Study Setting, Population, and Eligibility

The study was implemented in 14 health facilities located in Akwa Ibom State and Cross River State. The 2 states lie adjacent to each other in south-central Nigeria and share many sociocultural similarities. The population of interest was YLHIV aged 15 to 24 years who had been on ART for less than 12 months. In addition to age and ART eligibility criteria, participants had to demonstrate basic literacy for web-based chats, which was assessed at enrollment by asking the participant to read 3 short messages from intervention materials. We excluded individuals who were unable to attend the initial intervention group meeting if randomized to the intervention group, who were enrolled in another research study related to HIV retention or ART adherence, or who were severely ill and unable to provide informed consent at the time of recruitment. For IDIs, facilitators identified group members with high and low participation from whom participants were purposively selected to achieve representativeness by group participation, sex, and age.

Study Power

The study was powered to detect at least a 0.125 difference in the cumulative probability of retention at endline (0.45 in the control group and 0.575 in the intervention group), corresponding to a hazard ratio of 0.69, with 80% power and 5% significance level for a two-sided comparison using the log-rank test. This resulted in a total sample size of 500 (250 per study arm). Calculations also assumed exponential times to event and a 10% loss to study follow-up. Our estimate of retention was based on program data for the prior year from the United States Agency for International Development (USAID)–funded bilateral HIV services delivery (Strengthening Integrated Delivery of HIV/AIDS Services; SIDHAS) project, which supported HIV services in the facilities selected for this study.

Changes to the Study Protocol and Implementation

We originally planned to recruit YLHIV aged 15 to 22 years on ART for 6 months or less and expected to achieve our sample size within 3 months. Due to substantially lower than expected numbers of YLHIV enrolling in HIV treatment, we amended the study protocol and expanded the eligibility criteria to include YLHIV aged up to 24 years and on ART for 12 months or less. We also extended enrollment from 3 to 8 months and added 3 facilities (originally 11).

Finally, we intended to collect outcome data from participants after the intervention was completed and at 1 year from enrollment (approximately 6 months later) but were unable to collect a second round of outcome data because of prolonged recruitment—the funding project came to an end before the final round of data collection could take place.

Recruitment and Randomization

Eligible participants were sequentially recruited from patients attending clinic visits at the study facilities. Health care staff informed eligible participants of the study and directed those interested to a data collector who was stationed in the facility. Data collectors met with potential participants in a private setting in the facility, provided additional details on the study, then proceeded to obtain informed consent and enroll participants. Data collectors assigned participants to study arms using randomization envelopes. Randomization groups were concealed in sequentially numbered sealed opaque envelopes. The allocation sequence was generated using permuted blocks and stratified by local government areas (LGA), which are administrative subunits of the state, by a biostatistician otherwise uninvolved in the study using a validated SAS macro (SAS version 9.4) [49]. Recruitment occurred at health facilities, but randomization was managed from a central location for all facilities within an LGA. This was an open-label study with no blinding of the study staff or participants.

Measures

We collected demographic information (sex, age, relationship status, education, occupation, and religion) and information on HIV history, including date of HIV diagnosis, ART start date, disclosure of HIV status to others, and viral load testing. Additionally, we collected data on secondary outcomes of ART adherence, HIV-related knowledge, and social support as well as psychosocial factors associated with poor retention, including social isolation, depression, and perceived and experienced stigma.

Retention: For our primary time-to-event analysis, we computed the time retained in care from study enrollment to the date the participant was no longer classified as active on treatment, consistent with the President’s Emergency Plan for AIDS Relief (PEPFAR) indicator definitions [50]. We recorded dates for all scheduled clinic visits from patient medical records for each participant from study enrollment until the end of the study. If a participant failed to return after a scheduled visit for more than 28 days, the date of the missed visit was the date of loss to care recorded, unless death or transfer of service was documented before the missed visit. For a small number of patients, their first missed scheduled visit was scheduled on or within 28 days before study enrollment. If the participant missed this first scheduled visit by more than 28 days, he or she was assigned a retention time of 0.

We also reported on treatment status at endline. To be considered active on treatment at endline, an individual must have attended a visit or had a follow-up visit scheduled within 28 days of the date of their endline questionnaire. For those who did not complete an endline questionnaire, an approximate endline date was used (based on the median time in the study of those enrolled the same month). Participants who died or transferred facilities were categorized accordingly.

Measurement of secondary outcomes and other psychosocial variables are summarized in Table 1.

Data Collection

Structured questionnaires programmed into password-protected computer tablets were used to collect data from participants at both baseline and endline. We also extracted medical record data (MRE) from patient charts at both time points into a separate form programmed into the tablets. Data from tablets were uploaded daily to a secure computer server. Trained interviewers administered a semistructured IDI guide. IDIs were audio recorded and transcribed verbatim.

Data Analyses

All analyses used an intent-to-treat approach, in which all randomized participants were included in the statistical analysis and analyzed according to the group to which they were originally assigned, regardless of the treatment they received during the study. The primary hypothesis was that *YLHIV enrolled in HIV treatment services who participate in SMART Connections will be more likely to be retained in HIV care than YLHIV enrolled in HIV treatment services who do not participate in the intervention.*

Kaplan-Meier cumulative retention probabilities are reported with 95% CIs and plotted by study arm. Participants who were confirmed to have died or transferred to a facility outside the study facilities were considered censored. Participants who had elected to drop out of the study (n=8), or for whom all MRE were missing because of missing charts (n=14) or all visit data were missing data in charts (n=3), were excluded from the analysis.

The retention probabilities between the groups were compared using a log-rank test stratified by LGA with a two-sided α of .05. We also report on retention descriptively, examining lapses in care, and return to treatment over the course of study follow-up. The relationships between treatment exposure and secondary outcomes (ART adherence, HIV knowledge, and social support) as well as social isolation, depression, and HIV-related stigma were explored with *t* tests for continuous outcomes and chi-square test for categorical outcomes, using two-sided tests, with a significance level of 0.05.

We applied thematic analysis to the IDI data. A structured codebook was developed *a priori* based on the interview guide; emergent thematic codes were added during the analysis. To assess intercoder consistency, 2 analysts independently coded 3 transcripts, compared coding, and resolved differences through discussion. Coding was conducted using NVivo 12 [56]. Once all transcripts were coded, textual coding reports were produced.

Table 1. Secondary outcomes and related psychosocial variables and their measurement.

Concept	Measurement
HIV knowledge and treatment literacy	A set of 14 knowledge-based questions covering HIV transmission, diagnosis, treatment, and treatment monitoring based on topics covered in the SMART[a] Connections curriculum. Each item is scored 1 if answered correctly and 0 if answered incorrectly. A total knowledge score was calculated based on the proportion of items correctly answered.
Social support	Medical Outcomes Study-Social Support Survey, a 19-item scale that covers the dimensions of emotional, information, affectionate, and tangible social support in addition to positive social interaction [51]. Scores for each item were summed, then averaged. Range of possible scores: 0-5, with a higher score indicating a greater level of social support.
Adherence to antiretroviral treatment	Self-report using the AIDS Clinical Trials Group adherence questionnaire [52]. For these analyses, we limited our analysis to a single item that recorded self-reported ART[b] adherence for the 3 days before the interview. Values included 0 for no missed doses, and 1 for 1 or more missed doses.
Social isolation	4-item PROMIS[c] Social Isolation Scale [53]. Participants are asked how often they feel each of 4 different situations occurs in their lives. Responses options include 1=Never, 2=Rarely, 3=Sometimes, 4=Usually, and 5=Always. For this analysis, the total score was a sum of the responses for the 4 items.
Depression	Stanford Patient Education Research Center's PHQ-8[d] [54]. The PHQ-8 asks respondents on how many days over the prior 2 weeks they experienced 8 possible symptoms, with response options of *not at all*=0, *a few days*=1, *more than half the days*=2, and *most all of the days*=3. The score for each item is summed, and a total score that ranges from 0 to 24 is assigned. Respondents with scores <10 are classified as not depressed. Respondents who score 10-19 points are considered to have major depression, and those who score 20 or more have severe depression [54]. For these analyses, a categorical variable was created: 1=no depression, 2=major depression, and 3=severe depression.
HIV-related stigma	12-item scale adapted by Reinius and colleagues from the 40-item HIV stigma scale [55]. This shortened stigma scale covers 4 dimensions of stigma: personalized stigma, disclosure concerns, concerns about public attitudes, and negative self-image. Each item is scored on a 4-point Likert-type scale, and the scores added within dimensions with possible scores ranging from 3 to 12 [55]. A higher score indicates a greater level of perceived HIV-related stigma.

[a] SMART: Social Media to promote Adherence and Retention in Treatment.

[b] ART: antiretroviral therapy.

[c] PROMIS: Patient-Reported Outcomes Measurement Information System.

[d] PHQ-8: Patient Health Questionnaire Depression Scale.

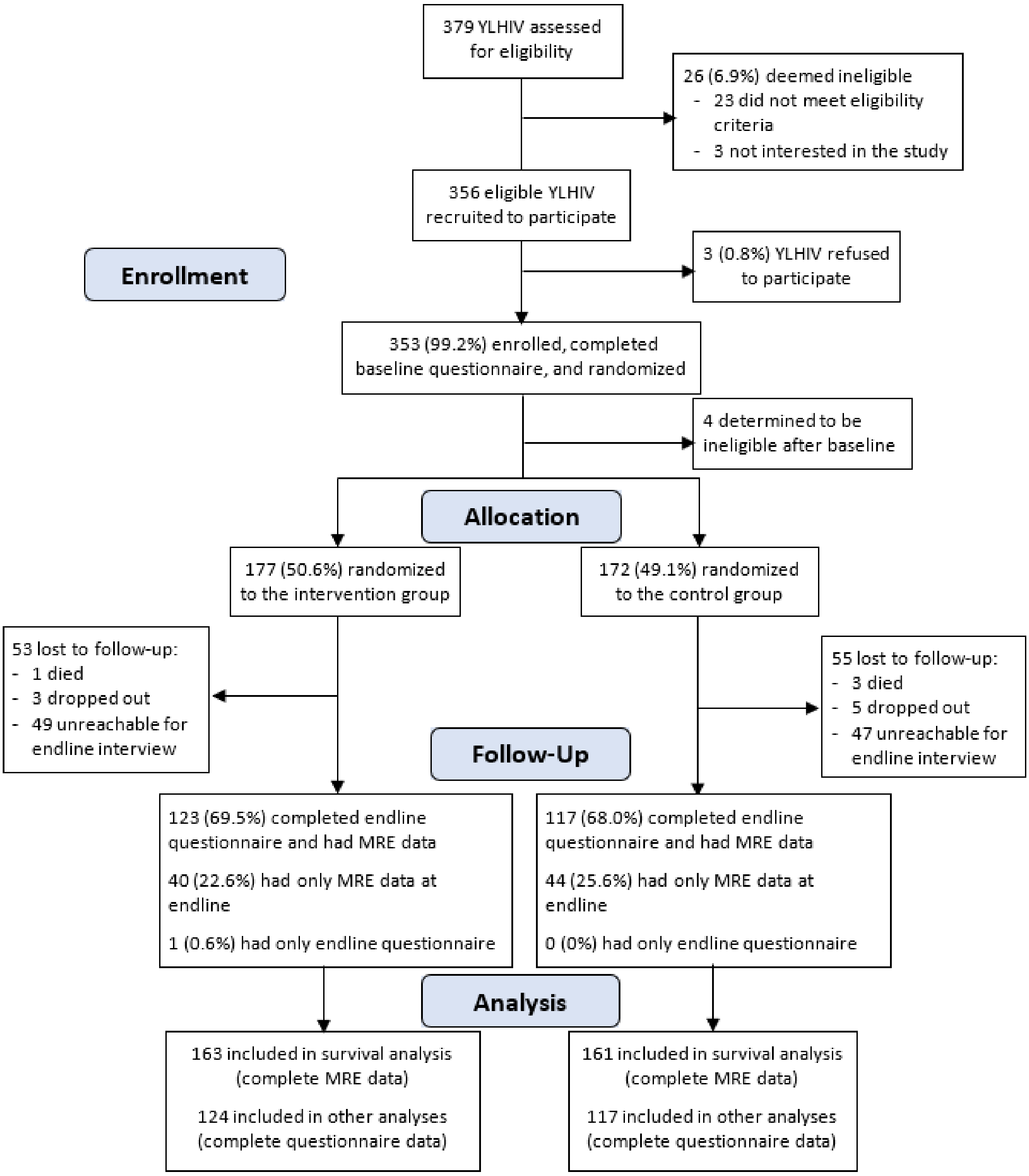
Results

Overview

We recruited 356 youths between September 2018 and April 2019, 353 (99.2%) of whom enrolled in the study (Figure 2). A total of 4 participants were removed after enrollment, as they provided false eligibility information, and were deemed ineligible. Participants were then randomly allocated to the intervention group (n=177) or control group (n=172). At endline (June to November 2019), 108 participants were lost to follow-up from the study, including 4 who died, 8 who discontinued study participation, and 96 who were not reachable for an endline interview. Endline interviews and MRE data were completed for 241 participants. MRE data were collected for an additional 84 participants who were not reachable for an endline interview but who had not elected to END study participation. We conducted IDIs with 21 intervention participants following the endline survey, of whom 16 were female and ranged in age from 17 to 24 years; 13 were classified as high participation and 8 as low participation.

A total of 10 randomization errors occurred during the study: 6 participants randomized to the control group were erroneously recorded by data collectors to be the intervention group and took part in the intervention; 4 participants randomized to the intervention group were erroneously recorded as in the control group and did not participate in the intervention. Two other participants, randomized to and recorded as in the control group, were erroneously contacted to participate in the intervention. In addition, 14 people randomized to the intervention group chose not to participate in the intervention. Thus, a total of 167 participants joined an online support group during the study: 160 randomized to the intervention group, 5 due to randomization errors, and 2 control group participants erroneously contacted.

Figure 2. Study flowchart. MRE: medical record data; YLHIV: youth living with HIV.



To assess the potential effects of loss to follow-up, we examined differences between those who did and did not complete an endline interview. There were no statistically significant differences in any background characteristic or baseline values of secondary outcomes between those who completed endline interviews and those lost to study follow-up (results not shown).

Background

Both study arms were similar in all baseline demographic characteristics (Table 2). Most participants were female 87.7% (306/349) and had completed some secondary school or more; the mean age was 21 years (SD 2.3). More than half of the respondents reported being married or in a relationship. Among those in a relationship (including married), fewer than half reported that their partner knew their HIV status, and about half knew their partner's HIV status. Most participants reported having access to a mobile phone in their household (303/349, 86.8%), with 67.1% (234/349) reporting that they owned their own mobile phone.

Table 2. Background characteristics of study participants at baseline (N=349).

Characteristic	Intervention (n=177)	Control (n=172)
Sex, n (%)		
Female	151 (85.3)	155 (90.1)
Male	26 (14.7)	17 (9.9)
Age (years), mean (SD)	21.3 (2.3)	21.0 (2.3)
Relationship status, n (%)[a]		
Married	37 (20.9)	40 (23.3)
Unmarried or in a relationship	83 (46.9)	85 (50.3)
Single	57 (32.2)	44 (26.0)
Among those married or in a relationship, n (%)		
Partner knows participant's HIV status[b]	42 (35.3)[c]	57 (45.6)[d]
Partner has HIV[e]	20 (16.7)[c]	26 (30.0)[d]
Partner does not have HIV	39 (32.5)[c]	40 (32.3)[d]
Does not know partner's HIV status	61 (50.8)[c]	58 (46.8)[d]
Education level, n (%)[e]		
Primary or less	13 (7.3)	30 (17.5)
Secondary	136 (76.8)	121 (70.8)
Any postsecondary	28 (15.8)	20 (11.7)
Currently working, n (%)	73 (41.2)	69 (40.1)
Religion, n (%)		
Protestant	152 (85.9)	150 (87.2)
Catholic	22 (12.4)	15 (8.7)
Other (all Christian denominations)	3 (1.7)	7 (4.1)
Time on ART[f] at enrollment (months), mean[g]	4.5	4.5
Pre-ART WHO[h] stage, n (%)[i]		
Stage 1	93 (60.8)	85 (57.4)
Stage 2	35 (22.9)	42 (28.4)
Stage 3	23 (15.0)	20 (13.5)
Stage 4	2 (1.3)	1 (0.7)
Had access to a phone in the home, n (%)	156 (88.1)	147 (85.5)
Owns a mobile phone, n (%)	118 (66.7)	116 (67.4)
Ever used social media sites of those who ever use the internet, n (%)		
Facebook	101 (57.0)	76 (44.2)
WhatsApp	69 (39.0)	52 (30.2)
Instagram	30 (16.9)	14 (8.1)
Snapchat	13 (7.3)	13 (7.6)
Other (Twitter, Tinder, Imo, etc)[b]	7 (4.0)	5 (2.9)

[a] 3 missing from the control group.

[b] 1 missing from the intervention group.

[c] n=120.

[d] n=125.

[e] 1 missing from the control group.

[f] ART: antiretroviral therapy.

[g] 15 missing from the intervention group, and 17 missing from the control group.

[h] WHO: World Health Organization.

[i] 24 missing in the control group, and 24 missing in the intervention group.

Retention in HIV Care

The probability of being retained in HIV care, defined as not having missed a scheduled appointment by more than 28 days, was similar between the intervention and control groups (Figure 3).

The probability of remaining in care without more than a 28-day gap was slightly higher in the intervention group than in the control group at each time point, except at 120 days; however, 95% CIs overlapped between the 2 study arms at all time points (Table 3).

Figure 3. Cumulative probability of retention in care (n=324).

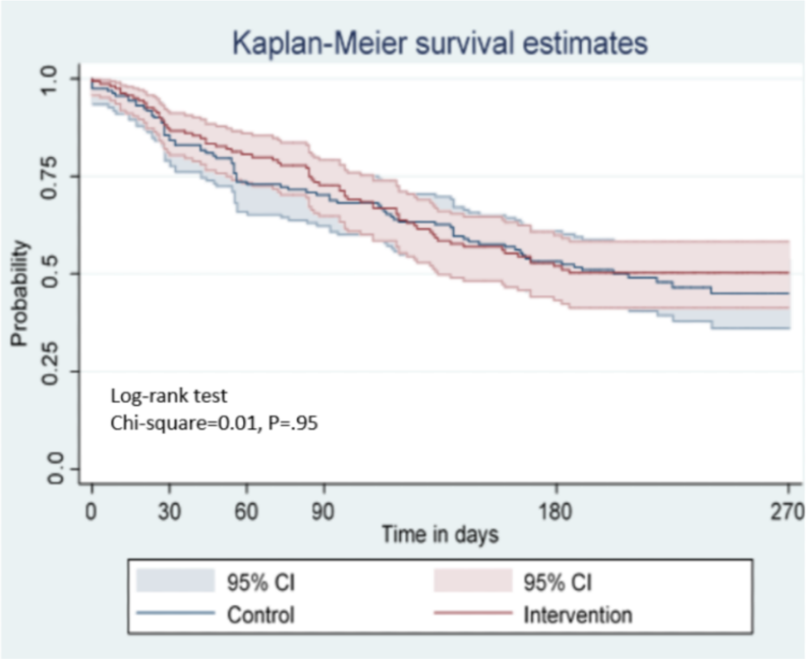


Table 3. Probabilities of remaining in care, without a gap of more than 28 days at 30, 60, 90, 180, and 270 days (n=324; 163 intervention and 161 control).

Time	Total at risk	Failures	Probability of being retained in care	SE	95% CI	
					Lower limit	Upper limit
Intervention						
0	163	1	0.99	0.01	0.96	1.00
30	129	21	0.87	0.03	0.80	0.91
60	116	9	0.81	0.03	0.73	0.86
90	102	11	0.73	0.04	0.65	0.79
180	59	27	0.52	0.04	0.43	0.60
270	8	2	0.50	0.04	0.41	0.58
Control						
0	161	4	0.98	0.01	0.94	0.99
30	132	25	0.84	0.03	0.78	0.89
60	108	17	0.73	0.04	0.65	0.79
90	103	4	0.70	0.04	0.62	0.77
180	72	24	0.53	0.04	0.45	0.61
270	12	8	0.45	0.04	0.36	0.54

Retention, measured as not having missed any scheduled visit by more than 28 days, did not reflect the participant's treatment status at endline (Table 4). Almost half of all participants missed at least one scheduled visit by more than 28 days during the 6 to 9 months of follow-up and were classified as lost to care at some point; however, most participants were active on treatment at endline. A total of 40.0% (65/163) in the treatment arm and 46.0% (74/161) in the control arm were counted as having been lost-to-care in the time-to-event analysis but were active on treatment at the end of the study.

Table 4. Clinic visits and retention over the study period for participants included in the retention analysis, intention to treat (n=324: 163 intervention and 161 control).

Characteristics	Intervention (n=163)	Control (n=161)
Number of clinic visits during study, mean (range, SD)	3.4 (0-10, 1.9)	3.9 (0-10, 1.9)
Missed at least one visit by >28 days, n (%)	70 (42.9)	78 (48.5)
Time between scheduled visits (months), n (%)[a]		
1	111 (72.1)	118 (75.6)
2	33 (21.4)	30 (19.2)
3	7 (4.5)	5 (3.2)
4	1 (0.6)	2 (1.3)
>5	2 (1.3)	1 (0.6)
Treatment status at endline, n (%)[b]		
Active on treatment	112 (75.7)	126 (83.4)
Lost to care	22 (14.9)	15 (9.9)
Transferred out	13 (8.8)	7 (4.6)
Deceased	1 (0.7)	3 (2.0)
Of those active on treatment at endline, n (%) who missed at least one scheduled visit by >28 days	45 (40.2)	58 (46.0)

[a] 9 missing from intervention group, and 5 missing from control group.

[b] 15 missing from intervention group, and 10 missing from control group.

The average number of clinic visits during the study varied by participant, in part because of variation in the length of time between scheduled visits. The average time between schedule visits was 1 month for most, but one-quarter of participants had an average time between visits ranging from 2 to more than five months.

Barriers to clinic attendance, as described in IDIs, included traveling away from home, anticipating stigma at the clinic, and competing obligations. When asked how easy it was to attend scheduled appointments, one participant described it as follows:

> *When you go to the hospital and they saw you…they will tell people about you, this is the kind of sickness that this person has, so that makes me very difficult to go.* [24-year-old female, high participation]

A few participants described hesitation leaving school, work, or family to attend appointments if it required disclosing their HIV status. Findings were similar in endline questionnaire open-ended data; among 79 participants who reported missing an appointment by one week or more, reasons for doing so included traveling (33/79, 42%), transportation issues (20/79, 25%), or were busy (13/79, 16.5%).

Most IDI participants described that the intervention impacted their attitudes and behaviors toward attending scheduled appointments. Some described that the intervention reinforced the importance of attending appointments to prevent gaps in medication and taught them to "take your drugs so that you'll feel okay " (21-year-old female, high participation). Others described learning strategies to avoid missing appointments, such as asking someone to collect refills if they could not. A few reported that the intervention served as a reminder to attend scheduled appointments; one participant mentioned receiving reminders from group members on their appointment day. Another described that being part of the group:

> *Something was there telling me to always don't miss appointment, to take my drugs, it was important, so I think the group was a reminder.* [19-year-old female, low participation]

In contrast, a few participants, most categorized as low group participation, felt that the intervention did not affect their retention in care.

Secondary Outcomes

HIV-related knowledge was high in both groups at endline but significantly higher in the intervention group compared with the control group (Table 5). No other statistically significant differences between study arms were observed for ART adherence, social support, social isolation, HIV-related stigma, or depression.

Table 5. Bivariate relationships between treatment group and endline HIV knowledge, psychosocial variables (*t* test), and self-reported adherence (chi-square test) among participants who responded to the endline questionnaire (n=241).

Outcomes	*t* value (*df*)	Chi-sqaure (*df*=1)	*P* value
HIV knowledge score	-2.96 (239)	N/A[a]	.003
Social isolation score	-0.79 (239)	N/A	.43
Social support score			
Total score[b]	-0.95 (238)	N/A	.34
Tangible subscore	-0.38 (239)	N/A	.70
Emotional/informational subscore[b]	-0.67 (238)	N/A	.51
Affectionate subscore[b]	-0.64 (238)	N/A	.53
Positive social interaction subscore	-1.57 (239)	N/A	.12
Stigma score			
Total score[c]	0.34 (196)	N/A	.73
Personalized stigma subscore[d]	-0.51 (209)	N/A	.61
Disclosure concerns subscore[e]	0.77 (235)	N/A	.44
Concerns about public attitudes subscore[f]	-0.54 (213)	N/A	.59
Negative self-image subscore[g]	0.02 (229)	N/A	.98
Depression[h]	N/A	0.15	.70
Adherence[b]	N/A	0.32	.57

[a] N/A: not applicable.

[b] 1 missing from the intervention group.

[c] 23 missing from the intervention group, 20 missing from the control group.

[d] 17 missing from the intervention group, 13 missing from the control group.

[e] 4 missing from the control group.

[f] 13 missing from the intervention group, 13 missing from the control group.

[g] 4 missing from the intervention group, 6 missing from the control group.

[h] 1 missing from the intervention group, 1 missing from the control group.

Although the 2 study arms did not differ significantly on social support, nearly all IDI participants, including those with low intervention participation, stated they received social support from facilitators and other group members. These participants described receiving encouragement and advice, having people to “share my feelings with” or “someone to talk to,” and receiving answers to factual and personal questions. One participant described the following:

> *I felt like I was not alone in the journey and it was really cool...it was amazing. I don't know how to say it in words but it's something to build (us up) because sometimes we can't just do it by ourselves. We need to find people in the situations with us for us get stronger, so the group actually made me stronger.* [19-year-old female, low participation]

Some support was related to self-management, such as encouragement to adhere to ART. For example, one participant described group members as:

> *…people that would encourage you no matter anything, they tell you no matter anything, that they’re okay with the drugs,… [they] encourage you to take the ART.* [22-year-old male, low participation]

Participants frequently received multiple types of support and often supported other group members by sharing their own experiences or providing emotional support and advice.

Nearly all IDI participants stated they felt a sense of connectedness with the group, sometimes described feeling as if they were “a family” or that participants got along “as brother and sister.” This feeling was often attributed to having group members of the same age range and HIV status and conferred a sense of safety and confidentiality within the group. In endline questionnaires, when asked what they liked most about the intervention, the most common responses included receiving encouragement and support, the ability to share their problems, and feeling a sense of unity or belonging.

When asked about sources of social support outside of the intervention group, most said they did not receive social support outside of the group, sometimes elaborating that they preferred not to disclose their HIV status to family and friends due to fear of stigma:

> *I don't like disclosing… I don’t know if that person is a victim of HIV…the person may start broadcasting me. Things like that so that’s why I don’t tell her, I didn’t tell people… They will not be able to encourage me since they don’t know what I’m passing through … I keep it [my status] to myself.* [19-year-old male, high participation]

Only a couple of IDI participants stated that they received support from family members; a few mentioned a health care provider.

Participant Perspectives on the Intervention

Nearly all who took part in the intervention (as treated) and completed an endline interview agreed that the intervention was useful to them, they enjoyed the intervention and felt comfortable interacting with the facilitators and other group members, and they would recommend the intervention to other YLHIV (Table 6). Nearly all also reported that connecting to the groups on Facebook was somewhat or very easy.

Table 6. Intervention participant perspectives on the web-based intervention at endline (n=127).

Characteristic	Intervention (n=127)[a]
Agree with the following statements, n (%)	
I enjoyed being a member of the online support group[b]	122 (97.6)
I received information during the support group that was useful to me[b]	124 (99.2)
Participating in the support group helped me better understand HIV infection[b]	124 (99.2)
I felt comfortable interacting with other group members[b]	116 (92.8)
I felt comfortable interacting with the group facilitator[c]	118 (95.2)
I made new friends in the group[b]	94 (75.2)
I would like to continue to be part of this group[b]	121 (96.8)
I think Facebook groups are a good way for young people on ART to interact with each other[b]	124 (99.2)
I think Facebook groups are a good way for support group leaders to get information to people on ART[d]	119 (99.2)
I would recommend this Facebook group to other young people living with HIV[b]	121 (96.8)
Ease of connecting to Facebook group, n (%)	
Very easy	80 (63.0)
Somewhat easy	32 (25.2)
Somewhat difficult	13 (10.2)
Very difficult	2 (1.6)

[a] This number includes the 6 participants assigned to the control group who took part in the intervention (as treated).

[b] 2 missing.

[c] 3 missing.

[d] 7 missing.

IDI participants often conveyed appreciation for learning about practical aspects of managing HIV, such as taking ART at consistent times during the day and eating “a balanced diet” to support overall health. Many participants also enjoyed learning why medication adherence is important. Participants recalled learning that ART adherence would help them feel healthier and achieve a longer life. One participant described as follows:

> *by taking the drugs and eating your food every day, your body will be okay … but if you avoid taking that drugs and you are not doing anything, you don’t go for test, you might die at any point in time and nobody will know the purpose of your death.* [22-year-old female, high participation]

Many IDI participants also expressed enthusiasm for the social and interactive elements of the intervention, such as riddles posed by the facilitator. Most felt the Facebook platform was acceptable, reporting that the “secret” groups ensured their privacy, and the web-based format allowed them to interact with the group at their convenience. One participant described as follows:

> *you’ll use it [the group] at the comfort of your home, not pressing you to go out, do this, every time to hear about the new information that you need to learn. Peacefully, you’ll just learn inside your room, inside your, the comfort of your own home. So, I was very happy about that one.* [17-year-old male, high participation]

Discussion

The SMART Connections intervention was designed to improve treatment retention among YLHIV by improving HIV-related knowledge and social support. Our findings indicate that the intervention did not significantly improve retention or social support; however, HIV-related knowledge did improve significantly. Data from IDIs provided evidence of perceived improvements in social support. Intervention participants also overwhelmingly found the intervention acceptable, liked the web-based platform, and reported it had helped them in their HIV treatment.

Our study adds to a small yet growing number of studies targeting YLHIV in LMICs. A 2019 literature review identified only 10 studies between 2016 and 2018 focused on adolescents (10-19 years) and/or YLHIV (15-24 years) in LMICs [36]. Among the 10 studies, 5 tested variations of youth-friendly services, 3 examined different community-based services to ALHIV/YLHIV, and 1 examined an SMS reminder system. Similar to our results, half of the studies reviewed—3 studies on youth-friendly services, 1 on community-based adherence clubs, and 1 on SMS reminders—showed no effects on retention [14]. Five studies found significant associations between youth-focused interventions and retention in HIV services; however, all were retrospective cohort studies, with limitations inherent to observational studies [14].

To date, most research on digital health interventions among PLHIV remains focused on high-income country settings, with a few exceptions [57-61]. A 2016 South African study examined the feasibility and acceptability of a web-based social media platform, MXit, to support YLHIV aged 12 to 25 years. Similar to our research, investigators found that the majority of youths in the study (84%) felt that offering a service outside of in-person meetings was useful [37]. A second South African study recently examined using mobile phones to provide peer mentorship for youths newly diagnosed with HIV [62]. This small case-control study also found no differences in 6- and 12-month retention or viral suppression between the groups [62]. Looking forward, interest in the use of digital health interventions to help meet the needs of YLHIV is growing. We identified published protocols of current studies examining interventions targeting youths; however, all 3 are being conducted in the United States [63-65].

SMART Connections was originally designed for adolescents aged 15 to 19 years living with HIV. The practical reality of ART initiation in our study settings, despite being among the highest HIV prevalence in Nigeria, led us to expand the age range for this study [66]. Most study participants were aged >19 years and although much of the content is relevant to the people of all ages, we believe future research should explore if some of the intervention content should be tailored to the differing developmental needs of those aged 15 to 19 years and those aged 20 to 24 years. We also noted that most (88%) of study participants were female. Although HIV prevalence among females in this age group is, on average, twice that of males in the age group, other research has demonstrated that the high ratio of females to males is not unusual in this setting [2]. A 2014 study examining the characteristics of YLHIV initiating ART in Nigeria found that from 2004 to 2013, 92% of new ART patients aged 15 to 24 years were female [42]. Given the high ratio of females to males, it may be worth exploring if and how the intervention could be better tailored to meet the gender-specific needs of young women.

Retention can be difficult to define and measure. Medical records, the source for our primary outcome, posed a particular problem as records contained errors, conflicting information, and missing data. In addition, most participants in care at endline had missed at least one appointment by more than 28 days, suggesting that, although missed visits and ART refills are problematic for effective treatment, they are a reality for many YLHIV, and missing visits does not necessarily translate to categorical loss to follow-up.

Measuring ART adherence is similarly difficult; self-reported gaps in adherence are underreported [67]. We originally considered using viral load as a marker for adherence; however, viral load test results were not widely available for participants, either because testing was not happening or results were not being recorded reliably. The study lacked sufficient resources to pay for testing for all participants. Future research should include viral load as a primary outcome for interventions designed to improve health outcomes among PLHIV.

Although we did not find a statistically significant intervention effect on social support, results from qualitative data indicate that participants developed friendships with, received support from, and grew close to other group participants despite interacting almost entirely on the web. One possible reason for the lack of an intervention effect on social support may have been that the measure we used was not HIV-specific; we may not have captured the type of social support participants felt they received through the intervention. Given challenges related to stigma and disclosure, further research is needed to operationalize and measure social support associated with HIV.

Limitations

Despite many strengths, including the use of rigorous experimental study design, the study had a number of limitations. Several factors limited the study implementation and perhaps the interpretation of results. First, the results may not be generalizable to other youths outside our study areas. In addition, recruitment lagged, taking more than twice as long as planned because YLHIV enrollment in ART was substantially lower than estimated using HIV service data before the study. Despite protocol modifications, we were only able to enroll 69.8% (349/500) of the planned participants, reducing the study’s power. Slow enrollment also prolonged the time necessary to enroll enough intervention participants to form a support group in many cases. Thus, participants contributed different amounts of time to the study, and some intervention group participants waited months before their support group could begin. This may have added to their risk of loss to follow up before initiating the intervention.

The context of HIV service delivery also changed during the study implementation. A dedicated *surge* in PEPFAR-supported HIV services aimed to increase the number of PLHIV enrolled on treatment and retained in care began in the study sites partway through the study [68]. This surge entailed intensive efforts to better support HIV treatment services and PLHIV in their care and treatment. Strategies including community-based

ART initiation, multimonth ART dispensing for stable patients, and community-delivered ART refills were implemented to increase ART initiation and to improve treatment adherence and retention [68]. These changes meant that the average time between clinic visits and the number of visits participants had during the study varied. Although participants in both study arms were exposed to this changing context, retention increased dramatically during the surge; retention at endline was considerably higher in both study arms compared with retention rates indicated by the presurge programmatic data used to calculate effect size estimates. These external efforts may have masked any possible impact that the intervention had on retention.

Conclusions

Our findings of improved HIV knowledge and high acceptability are encouraging despite a lack of measurable effect on retention. No single psychosocial intervention is likely to meet the varying needs of YLHIV, but the SMART Connections intervention appears to contribute to some of those needs. Digital health support groups may fill critical gaps in the services available for YLHIV. Given the increasing use of social media by youth, the platform may be a preferred resource for some aspects of HIV-related support. Web-based delivery of support group interventions through platforms such as Facebook can permit people to access information and other group members privately, when convenient, and without travel.

Moving forward, we suggest a few adaptations to the SMART Connections intervention and to continue to examine its potential effects. Expanding groups to include YLHIV on ART for more than 1 year may provide better support to those newly initiating treatment and help meet their own informational and support needs. Findings from the IDIs suggest that the intervention had a perceived effect on social support. We believe that the tool we used to measure social support may not have adequately distinguished between more general social support and social support related to HIV. Further work to develop and test measures of social support that better reflect the support given/provided in the context of HIV should be pursued. Finally, focusing on more reliable outcome measures, such as viral load, is strongly recommended for interventions attempting to improve HIV-related health outcomes.

Acknowledgments

The authors would like to acknowledge the contributions of several individuals to the successful implementation of the study. The USAID–funded SIDHAS project was essential for the successful implementation of the intervention under study. Drs Hadiza Khamofu and Tosin Idaboh of FHI 360 ensured the coordination of intervention implementation with the SIDHAS project. Dr Oluwasanmi Adedokun from FHI 360 provided key information and direct support for selecting study sites and informing sampling through the use of the SIDHAS program data. The authors would also like to thank Nnamdi Ike from FHI 360, who played a pivotal role in coordinating the logistics for study-related training and liaising between the health facilities, the study team, and the local partners. Mr Ike also was key to coordinating efforts with the community-based organizations and monitoring the implementation of the intervention under study.

The authors would like to acknowledge the Center for Population and Reproductive Health study team and their efforts to successfully carry out data collection and enroll study participants in the intervention. The authors especially would also like to thank Ms Hannah Silverstein from the University of North Carolina for her creative efforts in illustrating the intervention and contributions to content development. Finally, the authors would like to thank the facilitators and youths who took part in the study, without whom this research would not have been possible.

This study was funded by the USAID through the US PEPFAR, under task order contract number AID-OAA-TO-15-00003 and YouthPower Action, under Indefinite Delivery, Indefinite Quantity contract number AID-OAA-I-15-00009, YouthPower: Implementation. The contents of this publication are the sole responsibility of FHI 360 and do not necessarily reflect the views of the USAID or the US Government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention implementation guide.
[PDF File (Adobe PDF File), 5352 KB - jmir_v22i6e18343_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1556 KB - jmir_v22i6e18343_app2.pdf]

References

1. World Health Organization. 2019. Adolescents: Health Risks and Solutions URL: https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions [accessed 2020-04-29]

2. UNAIDS. 2019. UNAIDS Data URL: https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf [accessed 2020-04-29]
3. UNAIDS. 2013. Global Report: UNAIDS Report on the Global AIDS Epidemic 2013 URL: http://files.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf [accessed 2020-04-29]
4. World Health Organization. Adolescent Health Epidemiology URL: https://www.who.int/maternal_child_adolescent/epidemiology/adolescence/en/ [accessed 2020-02-13]
5. UNAIDS. 2016. Ending The AIDS Epidemic For Adolescents, With Adolescents: A Practical Guide to Meaningfully Engage Adolescents In the AIDS Response URL: https://www.unaids.org/sites/default/files/media_asset/ending-AIDS-epidemic-adolescents_en.pdf [accessed 2020-04-29]
6. UNICEF. For Every Child, End AIDS: Seventh Stocktaking Report, 2016. New York, USA: UNICEF; 2016.
7. Koenig LJ, Nesheim S, Abramowitz S. Adolescents with perinatally acquired HIV: emerging behavioral and health needs for long-term survivors. Curr Opin Obstet Gynecol 2011 Oct;23(5):321-327. [doi: 10.1097/GCO.0b013e32834a581b] [Medline: 21836510]
8. Lamb MR, Fayorsey R, Nuwagaba-Biribonwoha H, Viola V, Mutabazi V, Alwar T, et al. High attrition before and after ART initiation among youth (15-24 years of age) enrolled in HIV care. AIDS 2014 Feb 20;28(4):559-568 [FREE Full text] [doi: 10.1097/QAD.0000000000000054] [Medline: 24076661]
9. Adejumo OA, Malee KM, Ryscavage P, Hunter SJ, Taiwo BO. Contemporary issues on the epidemiology and antiretroviral adherence of HIV-infected adolescents in sub-saharan Africa: a narrative review. J Int AIDS Soc 2015;18:20049 [FREE Full text] [doi: 10.7448/IAS.18.1.20049] [Medline: 26385853]
10. Ojwang' VO, Penner J, Blat C, Agot K, Bukusi EA, Cohen CR. Loss to follow-up among youth accessing outpatient HIV care and treatment services in Kisumu, Kenya. AIDS Care 2016;28(4):500-507 [FREE Full text] [doi: 10.1080/09540121.2015.1110234] [Medline: 26565428]
11. UNAIDS. 2014. Fast-Track - Ending the AIDS epidemic by 2030 URL: https://www.unaids.org/en/resources/documents/2014/JC2686_WAD2014report [accessed 2020-04-29]
12. Murray KR, Dulli LS, Ridgeway K, dal Santo L, de Mora DD, Olsen P, et al. Improving retention in HIV care among adolescents and adults in low- and middle-income countries: a systematic review of the literature. PLoS One 2017;12(9):e0184879 [FREE Full text] [doi: 10.1371/journal.pone.0184879] [Medline: 28961253]
13. Ridgeway K, Dulli LS, Murray KR, Silverstein H, dal Santo L, Olsen P, et al. Interventions to improve antiretroviral therapy adherence among adolescents in low- and middle-income countries: a systematic review of the literature. PLoS One 2018;13(1):e0189770 [FREE Full text] [doi: 10.1371/journal.pone.0189770] [Medline: 29293523]
14. Casale M, Carlqvist A, Cluver L. Recent interventions to improve retention in HIV care and adherence to antiretroviral treatment among adolescents and youth: a systematic review. AIDS Patient Care STDS 2019 Jun;33(6):237-252. [doi: 10.1089/apc.2018.0320] [Medline: 31166783]
15. MacKenzie RK, van Lettow M, Gondwe C, Nyirongo J, Singano V, Banda V, et al. Greater retention in care among adolescents on antiretroviral treatment accessing 'teen club' an adolescent-centred differentiated care model compared with standard of care: a nested case-control study at a tertiary referral hospital in Malawi. J Int AIDS Soc 2017 Nov;20(3):e25028 [FREE Full text] [doi: 10.1002/jia2.25028] [Medline: 29178197]
16. McBride K, Parent J, Mmanga K, Chivwala M, Nyirenda MH, Schooley A, et al. ART adherence among Malawian youth enrolled in teen clubs: a retrospective chart review. AIDS Behav 2019 Sep;23(9):2629-2633. [doi: 10.1007/s10461-019-02580-y] [Medline: 31292826]
17. Zanoni BC, Sibaya T, Cairns C, Lammert S, Haberer JE. Higher retention and viral suppression with adolescent-focused HIV clinic in South Africa. PLoS One 2017;12(12):e0190260 [FREE Full text] [doi: 10.1371/journal.pone.0190260] [Medline: 29287088]
18. Fatti G, Jackson D, Goga AE, Shaikh N, Eley B, Nachega JB, et al. The effectiveness and cost-effectiveness of community-based support for adolescents receiving antiretroviral treatment: an operational research study in South Africa. J Int AIDS Soc 2018 Feb;21(Suppl 1) [FREE Full text] [doi: 10.1002/jia2.25041] [Medline: 29485714]
19. Hickey MD, Salmen CR, Omollo D, Mattah B, Fiorella KJ, Geng EH, et al. Implementation and operational research: pulling the network together: quasiexperimental trial of a patient-defined support network intervention for promoting engagement in HIV care and medication adherence on Mfangano island, Kenya. J Acquir Immune Defic Syndr 2015 Aug 1;69(4):e127-e134 [FREE Full text] [doi: 10.1097/QAI.0000000000000664] [Medline: 25984711]
20. Muñoz M, Bayona J, Sanchez E, Arevalo J, Sebastian JL, Arteaga F, et al. Matching social support to individual needs: a community-based intervention to improve HIV treatment adherence in a resource-poor setting. AIDS Behav 2011 Oct;15(7):1454-1464 [FREE Full text] [doi: 10.1007/s10461-010-9697-9] [Medline: 20383572]
21. Achieng L, Musangi H, Ong'uti S, Ombegoh E, Bryant L, Mwiindi J, et al. An observational cohort comparison of facilitators of retention in care and adherence to anti-eetroviral therapy at an HIV treatment center in Kenya. PLoS One 2012;7(3):e32727 [FREE Full text] [doi: 10.1371/journal.pone.0032727] [Medline: 22427869]
22. Bateganya MH, Amanyeiwe U, Roxo U, Dong M. Impact of support groups for people living with HIV on clinical outcomes: a systematic review of the literature. J Acquir Immune Defic Syndr 2015 Apr 15;68(Suppl 3):S368-S374 [FREE Full text] [doi: 10.1097/QAI.0000000000000519] [Medline: 25768876]

23. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
24. Orrell C, Cohen K, Mauff K, Bangsberg DR, Maartens G, Wood R. A randomized controlled trial of real-time electronic adherence monitoring with text message dosing reminders in people starting first-line antiretroviral therapy. J Acquir Immune Defic Syndr 2015 Dec 15;70(5):495-502. [doi: 10.1097/QAI.0000000000000770] [Medline: 26218411]
25. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
26. Rodrigues R, Shet A, Antony J, Sidney K, Arumugam K, Krishnamurthy S, et al. Supporting adherence to antiretroviral therapy with mobile phone reminders: results from a cohort in South India. PLoS One 2012;7(8):e40723 [FREE Full text] [doi: 10.1371/journal.pone.0040723] [Medline: 22952574]
27. Sabin LL, DeSilva MB, Gill CJ, Zhong L, Vian T, Xie W, et al. Improving adherence to antiretroviral therapy with triggered real-time text message reminders: the China adherence through technology study. J Acquir Immune Defic Syndr 2015 Aug 15;69(5):551-559 [FREE Full text] [doi: 10.1097/QAI.0000000000000651] [Medline: 25886927]
28. Uzma Q, Emmanuel F, Ather U, Zaman S. Efficacy of interventions for improving antiretroviral therapy adherence in HIV/AIDS cases at PIMS, Islamabad. J Int Assoc Physicians AIDS Care (Chic) 2011;10(6):373-383. [doi: 10.1177/1545109710383175] [Medline: 21317163]
29. Elul B, Lamb MR, Lahuerta M, Abacassamo F, Ahoua L, Kujawski SA, et al. A combination intervention strategy to improve linkage to and retention in HIV care following diagnosis in Mozambique: a cluster-randomized study. PLoS Med 2017 Nov;14(11):e1002433 [FREE Full text] [doi: 10.1371/journal.pmed.1002433] [Medline: 29136001]
30. Abdulrahman SA, Rampal L, Ibrahim F, Radhakrishnan AP, Shahar HK, Othman N. Mobile phone reminders and peer counseling improve adherence and treatment outcomes of patients on ART in Malaysia: a randomized clinical trial. PLoS One 2017;12(5):e0177698 [FREE Full text] [doi: 10.1371/journal.pone.0177698] [Medline: 28520768]
31. Nsagha DS, Lange I, Fon PN, Assob JC, Tanue EA. A randomized controlled trial on the usefulness of mobile text phone messages to improve the quality of care of HIV and AIDS patients in Cameroon. Open AIDS J 2016;10:93-103 [FREE Full text] [doi: 10.2174/1874613601610010093] [Medline: 27583062]
32. Belzer ME, MacDonell KK, Clark LF, Huang J, Olson J, Kahana SY, Adolescent Medicine Trials Network for HIV/AIDS Interventions. Acceptability and feasibility of a cell phone support intervention for youth living with HIV with nonadherence to antiretroviral therapy. AIDS Patient Care STDS 2015 Jun;29(6):338-345 [FREE Full text] [doi: 10.1089/apc.2014.0282] [Medline: 25928772]
33. Dowshen N, Kuhns LM, Gray C, Lee S, Garofalo R. Feasibility of interactive text message response (ITR) as a novel, real-time measure of adherence to antiretroviral therapy for HIV+ youth. AIDS Behav 2013 Jul;17(6):2237-2243. [doi: 10.1007/s10461-013-0464-6] [Medline: 23546844]
34. Dowshen N, Kuhns LM, Johnson A, Holoyda BJ, Garofalo R. Improving adherence to antiretroviral therapy for youth living with HIV/AIDS: a pilot study using personalized, interactive, daily text message reminders. J Med Internet Res 2012 Apr 5;14(2):e51 [FREE Full text] [doi: 10.2196/jmir.2015] [Medline: 22481246]
35. Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A randomized controlled trial of personalized text message reminders to promote medication adherence among HIV-positive adolescents and young adults. AIDS Behav 2016 May;20(5):1049-1059 [FREE Full text] [doi: 10.1007/s10461-015-1192-x] [Medline: 26362167]
36. Gaysynsky A, Romansky-Poulin K, Arpadi S. 'My YAP family': analysis of a Facebook group for young adults living with HIV. AIDS Behav 2015 Jun;19(6):947-962. [doi: 10.1007/s10461-014-0887-8] [Medline: 25186783]
37. Henwood R, Patten G, Barnett W, Hwang B, Metcalf C, Hacking D, et al. Acceptability and use of a virtual support group for HIV-positive youth in Khayelitsha, Cape Town using the MXit social networking platform. AIDS Care 2016 Jul;28(7):898-903. [doi: 10.1080/09540121.2016.1173638] [Medline: 27098208]
38. MacPherson P, Munthali C, Ferguson J, Armstrong A, Kranzer K, Ferrand RA, et al. Service delivery interventions to improve adolescents' linkage, retention and adherence to antiretroviral therapy and HIV care. Trop Med Int Health 2015 Aug;20(8):1015-1032 [FREE Full text] [doi: 10.1111/tmi.12517] [Medline: 25877007]
39. Enane LA, Davies M, Leroy V, Edmonds A, Apondi E, Adedimeji A, et al. Traversing the cascade: urgent research priorities for implementing the 'treat all' strategy for children and adolescents living with HIV in sub-Saharan Africa. J Virus Erad 2018 Nov 15;4(Suppl 2):40-46 [FREE Full text] [Medline: 30515313]
40. World Health Organization. 2017. A Global Research Agenda for Adolescents Living With HIV: Research for an AIDS Free Generation URL: https://www.who.int/hiv/pub/toolkits/cipher-research-adolescents-living-with-hiv/en/ [accessed 2020-04-29]
41. US Embassy and Consulate in Nigeria. 2016. Nigeria Country Operational Plan (COP) 2016 Strategic Direction Summary URL: https://ng.usembassy.gov/wp-content/uploads/sites/177/2017/02/2016_Strategic_Direction_Summary_Final_June-6-2016.pdf [accessed 2020-04-29]
42. Auld AF, Agolory SG, Shiraishi RW, Wabwire-Mangen F, Kwesigabo G, Mulenga M, et al. Antiretroviral therapy enrollment characteristics and outcomes among HIV-infected adolescents and young adults compared with older adults-seven African

countries, 2004-2013. MMWR Morb Mortal Wkly Rep 2014 Nov 28;63(47):1097-1103 [FREE Full text] [Medline: 25426651]

43. Poushter J, Oates R. Pew Research Center. 2015. Cell Phones in Africa: Communication Lifeline URL: http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/ [accessed 2020-02-15]
44. Akinfaderin-Agarau F, Chirtau M, Ekponimo S, Power S. Opportunities and limitations for using new media and mobile phones to expand access to sexual and reproductive health information and services for adolescent girls and young women in six Nigerian states. Afr J Reprod Health 2012 Jun;16(2):219-230. [Medline: 22916554]
45. Dulli L, Ridgeway K, Packer C, Plourde KF, Mumuni T, Idaboh T, et al. An online support group intervention for adolescents living with HIV in Nigeria: a pre-post test study. JMIR Public Health Surveill 2018 Nov 28;4(4):e12397 [FREE Full text] [doi: 10.2196/12397] [Medline: 30487116]
46. Cohen S, Underwood LG, Gottlieb BH, editors. Social Support Measurement and Intervention: A Guide for Health and Social Scientists. New York, USA: Oxford University Press; 2000.
47. Heaney C, Israel B. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2008.
48. FHI 360. 2013. Positive Connections: Leading Information and Support Groups for Adolescents Living with HIV URL: https://www.fhi360.org/resource/positive-connections-leading-information-and-support-groups-adolescents-living-hiv [accessed 2020-04-29]
49. SAS: Analytics, Artificial Intelligence and Data Management. 2019. SAS 9.4 URL: https://www.sas.com/en_ca/software/sas9.html [accessed 2020-04-29]
50. United States Department of State. 2018. Monitoring, Evaluation, and Reporting Indicator Reference Guide URL: https://www.state.gov/wp-content/uploads/2019/10/PEPFAR-MER-Indicator-Reference-Guide-Version-2.4-FY20.pdf [accessed 2020-02-15]
51. Sherbourne C, Stewart A. The MOS social support survey. Soc Sci Med 1991;32(6):705-714. [doi: 10.1016/0277-9536(91)90150-b] [Medline: 2035047]
52. Chesney MA, Ickovics JR, Chambers DB, Gifford AL, Neidig J, Zwickl B, et al. Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG adherence instruments. Patient care committee & adherence working group of the outcomes committee of the adult AIDS clinical trials group (AACTG). AIDS Care 2000 Jun;12(3):255-266. [doi: 10.1080/09540120050042891] [Medline: 10928201]
53. Health Measures. 2015. Social Isolation: A Brief Guide to the PROMIS Social Isolation Instruments URL: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Social_Isolation_Scoring_Manual.pdf [accessed 2020-02-15]
54. Self Management Research Center. 2013. Personal Health Questionnaire Depression Scale (PHQ-8) URL: https://www.selfmanagementresource.com/docs/pdfs/English_-_phq.pdf [accessed 2020-02-15]
55. Reinius M, Wettergren L, Wiklander M, Svedhem V, Ekström AM, Eriksson LE. Development of a 12-item short version of the HIV stigma scale. Health Qual Life Outcomes 2017 May 30;15(1):115 [FREE Full text] [doi: 10.1186/s12955-017-0691-z] [Medline: 28558805]
56. QSR International. 2019. NVivo Product 12th Editions URL: https://help-nv.qsrinternational.com/12/win/v12.1.90-d3ea61/Content/about-nvivo/product-editions.htm [accessed 2020-04-29]
57. Rhodes SD, McCoy TP, Tanner AE, Stowers J, Bachmann LH, Nguyen AL, et al. Using social media to increase HIV testing among gay and bisexual men, other men who have sex with men, and transgender persons: outcomes from a randomized community trial. Clin Infect Dis 2016 Jun 1;62(11):1450-1453 [FREE Full text] [doi: 10.1093/cid/ciw127] [Medline: 26980878]
58. Tanner AE, Mann L, Song E, Alonzo J, Schafer K, Arellano E, et al. weCARE: a social media-based intervention designed to increase HIV care linkage, retention, and health outcomes for racially and ethnically diverse young MSM. AIDS Educ Prev 2016 Jun;28(3):216-230 [FREE Full text] [doi: 10.1521/aeap.2016.28.3.216] [Medline: 27244190]
59. Iribarren SJ, Ghazzawi A, Sheinfil AZ, Frasca T, Brown W, Lopez-Rios J, et al. Mixed-method evaluation of social media-based tools and traditional strategies to recruit high-risk and hard-to-reach populations into an HIV prevention intervention study. AIDS Behav 2018 Jan;22(1):347-357 [FREE Full text] [doi: 10.1007/s10461-017-1956-6] [Medline: 29124420]
60. Tanner AE, Song EY, Mann-Jackson L, Alonzo J, Schafer K, Ware S, et al. Preliminary impact of the weCARE social media intervention to support health for young men who have sex with men and transgender women with HIV. AIDS Patient Care STDS 2018 Nov;32(11):450-458 [FREE Full text] [doi: 10.1089/apc.2018.0060] [Medline: 30398955]
61. Bauermeister JA, Muessig KE, LeGrand S, Flores DD, Choi SK, Dong W, et al. HIV and sexuality stigma reduction through engagement in online forums: results from the HealthMpowerment intervention. AIDS Behav 2019 Mar;23(3):742-752 [FREE Full text] [doi: 10.1007/s10461-018-2256-5] [Medline: 30121727]
62. Hacking D, Mgengwana-Mbakaza Z, Cassidy T, Runeyi P, Duran LT, Mathys RH, et al. Peer mentorship via mobile phones for newly diagnosed HIV-positive youths in clinic care in Khayelitsha, South Africa: mixed methods study. J Med Internet Res 2019 Dec 10;21(12):e14012 [FREE Full text] [doi: 10.2196/14012] [Medline: 31821150]

63. Arnold EM, Swendeman D, Harris D, Fournier J, Kozina L, Abdalian S, Adolescent Medicine Trials Network CARES Team. The stepped care intervention to suppress viral load in youth living with HIV: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Feb 27;8(2):e10791 [FREE Full text] [doi: 10.2196/10791] [Medline: 30810536]
64. Medich M, Swendeman DT, Comulada WS, Kao UH, Myers JJ, Brooks RA, Special Projects Of National Significance Social Media Initiative Study Group. Promising approaches for engaging youth and young adults living with HIV in HIV primary care using social media and mobile technology interventions: protocol for the SPNS social media initiative. JMIR Res Protoc 2019 Jan 31;8(1):e10681 [FREE Full text] [doi: 10.2196/10681] [Medline: 30702434]
65. Swendeman D, Arnold EM, Harris D, Fournier J, Comulada WS, Reback C, Adolescent Medicine Trials Network (ATN) CARES Team. Text-messaging, online peer support group, and coaching strategies to optimize the HIV prevention continuum for youth: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Aug 9;8(8):e11165 [FREE Full text] [doi: 10.2196/11165] [Medline: 31400109]
66. Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS). 2019. National Summary Sheet - Preliminary Findings URL: https://www.naiis.ng/resource/factsheet/NAIIS%20PA%20NATIONAL%20FACTSHEET%20FINAL.pdf [accessed 2020-04-28]
67. Iacob SA, Iacob DG, Jugulete G. Improving the adherence to antiretroviral therapy, a difficult but essential task for a successful HIV treatment-clinical points of view and practical considerations. Front Pharmacol 2017;8:831 [FREE Full text] [doi: 10.3389/fphar.2017.00831] [Medline: 29218008]
68. FHI 360. 2019. Implementing the Surge HIV Response in Akwa Ibom: An Accelerated HIV Epidemic Control Drive URL: https://www.fhi360.org/sites/default/files/media/documents/resource-nigeria-hiv-surge-hiv-response.pdf [accessed 2020-04-29]

Abbreviations

ART: antiretroviral therapy
IDI: in-depth interview
LGA: local government area
LMIC: low- and middle-income country
MRE: medical record data
PEPFAR: President’s Emergency Plan for AIDS Relief
PLHIV: people living with HIV
SIDHAS: Strengthening Integrated Delivery of HIV/AIDS Services
SMART: Social Media to promote Adherence and Retention in Treatment
SSA: sub-Saharan Africa
USAID: United States Agency for International Development
YLHIV: youth living with HIV

Edited by G Eysenbach; submitted 24.02.20; peer-reviewed by DM Hardey, R Mpofu, C Carrion; comments to author 16.03.20; revised version received 25.03.20; accepted 10.04.20; published 02.06.20.

Please cite as:
Dulli L, Ridgeway K, Packer C, Murray KR, Mumuni T, Plourde KF, Chen M, Olumide A, Ojengbede O, McCarraher DR
A Social Media–Based Support Group for Youth Living With HIV in Nigeria (SMART Connections): Randomized Controlled Trial
J Med Internet Res 2020;22(6):e18343
URL: https://www.jmir.org/2020/6/e18343
doi:10.2196/18343
PMID:32484444

Review

The Effectiveness of Digital Health Interventions in the Management of Musculoskeletal Conditions: Systematic Literature Review

Stephanie Hewitt[1*], MSc; Ruth Sephton[1*], MSc; Gillian Yeowell[2*], PhD

[1]St Helens Therapy Department, North West Boroughs Healthcare NHS Foundation Trust, Merseyside, United Kingdom

[2]Department of Health Professions, Manchester Metropolitan University, Manchester, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Gillian Yeowell, PhD
Department of Health Professions
Manchester Metropolitan University
34 Bonsall Street
Manchester, M15 6GX
United Kingdom
Phone: 44 1612472961
Email: g.yeowell@mmu.ac.uk

Abstract

Background: Musculoskeletal conditions are the second greatest contributor to disability worldwide and have significant individual, societal, and economic implications. Due to the growing burden of musculoskeletal disability, an integrated and strategic response is urgently required. Digital health interventions provide high-reach, low-cost, readily accessible, and scalable interventions for large patient populations that address time and resource constraints.

Objective: This review aimed to investigate if digital health interventions are effective in reducing pain and functional disability in patients with musculoskeletal conditions.

Methods: A systematic review was undertaken to address the research objective. The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. The review protocol was registered with the International Prospective Register of Systematic Reviews before commencement of the study. The following databases were searched: Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica database (EMBASE), Cumulative Index to Nursing and Allied Health Literature, and Scopus from January 1, 2000, to November 15, 2019, using search terms and database specific−medical subject headings terms in various combinations appropriate to the research objective.

Results: A total of 19 English language studies were eligible for inclusion. Of the 19 studies that assessed musculoskeletal pain, 9 reported statistically significant reductions following digital intervention. In all, 16 studies investigated functional disability; 10 studies showed a statistically significant improvement. Significant improvements were also found in a range of additional outcomes. Due to the heterogeneity of the results, a meta-analysis was not feasible.

Conclusions: This review has demonstrated that digital health interventions have some clinical benefits in the management of musculoskeletal conditions for pain and functional disability. Digital health interventions have the potential to contribute positively toward reducing the multifaceted burden of musculoskeletal conditions to the individual, economy, and society.

Trial Registration: PROSPERO CRD42018093343; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=93343

(J Med Internet Res 2020;22(6):e15617) doi:10.2196/15617

KEYWORDS

musculoskeletal pain; physical functional performance; health communication; online intervention; web-based intervention; mobile phone

Introduction

Background

Musculoskeletal conditions are the second greatest contributor to disability worldwide and have substantial individual, societal, and economic implications [1,2]. The term musculoskeletal conditions is a broad term used to describe a large number of conditions that affect bones, joints, and soft tissues [3]. Musculoskeletal conditions comprise over 100 different disorders, diseases, and syndromes, most of which affect people's ability to carry out normal activities and impact their quality of life. The most prevalent of these conditions are low back pain (LBP) and osteoarthritis (OA) [3]. Musculoskeletal conditions account for 30% of general practitioner consultations in England and are associated with related comorbidities, including diabetes, depression, and obesity [4]. For the individual, the most common symptoms include pain, aching, stiffness, fatigue, reduced physical functioning, and loss of dexterity [3]. Treating musculoskeletal conditions is estimated to cost the United States US $213 billion [5] and costs the UK economy £10.2 billion (US $12.62 billion) in direct costs to the National Health Service [3].

Due to the prevalence and growing burden of musculoskeletal disability, an integrated, strategic approach that provides effective and accessible models of health service delivery on a population level is urgently required [2,6]. The use of mobile and wireless digital health interventions is one possible solution to deliver this objective [7]. Digital health interventions provide opportunities to tackle health system challenges and offer the potential to enhance the quality and sustainability of musculoskeletal health services [8]. The World Health Organization has recently published guidelines that classify digital health interventions in an attempt to standardize the vocabulary used within the diverse communities working in digital health [8]. In cognizance of this, digital health interventions in this study applies to interventions for clients, with all digital health interventions being delivered as apps, via websites or via web-based software [8].

Digital health interventions can provide high-reach, low-cost, readily accessible, and scalable patient education and self-management interventions that address time and resource constraints for musculoskeletal populations, delivered via apps or web-based platforms [6]. However, there are problems facing the implementation of digital health interventions [9]. A common problem is the failure of agreement on what constitutes appropriate evaluation before widespread rollout [9,10]. In addition, there is tension between the dynamic development of digital interventions and the slow transition into clinical practice from more conventional clinical trial outcomes [9-11]. If digital interventions are to be utilized as therapeutic interventions for musculoskeletal conditions, clinicians and service users need to have confidence in their effectiveness [10,11]. Ultimately, the development and utilization of digital health interventions in a therapeutic capacity for musculoskeletal conditions need to work toward reducing the burden of musculoskeletal-related disability.

Objective

The primary aim of this systematic review was to assess if digital health interventions are clinically effective in reducing pain and functional disability in patients with musculoskeletal conditions. The secondary aim was to explore the content, characteristics, and delivery of digital health interventions in the studies identified to ascertain if there were specific aspects of the interventions, or the population they were targeting, that were associated with beneficial outcomes. To the best of our knowledge, no systematic review has examined the effectiveness of digital health interventions in the management of musculoskeletal conditions.

Methods

Overview

A systematic review was undertaken to address the aims of the study. The authors (SH and RS) were assisted in the literature search by an experienced librarian, proficient in searching medical databases. The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [12]. The review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO reference: CRD42018093343) before commencement of the study. However, deviations from the protocol were required following the pilot study. This was related to the inclusion of the PsycINFO database, which yielded no useful results and, as such, was not included in the main search.

Search Strategy

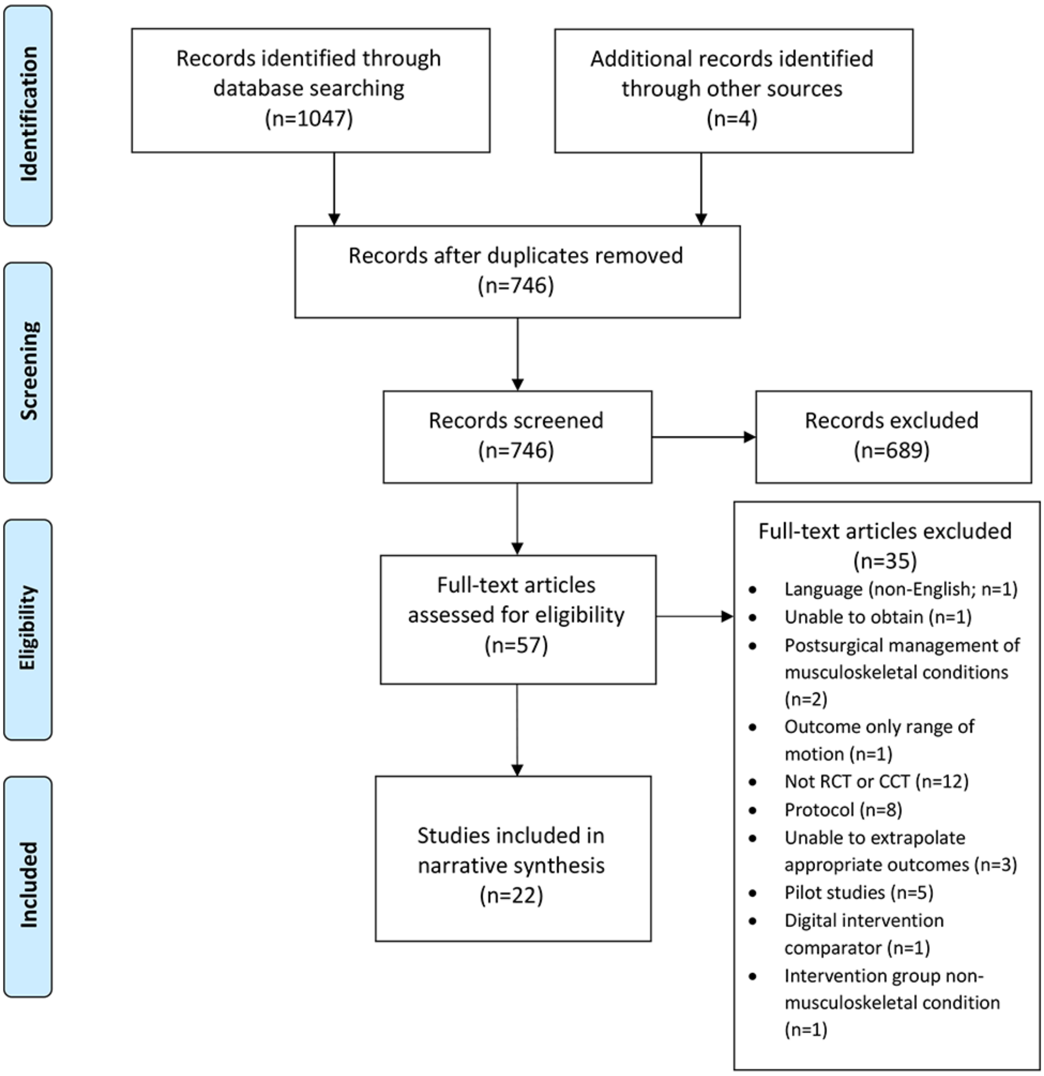
A systematic search of the following databases was conducted: MEDLINE, EMBASE, Cumulative Index to Nursing and Allied Health Literature, and Scopus from January 1, 2000 to November 15, 2019. The search was conducted on November 15, 2019. Abstract and subject-heading search terms pertinent to the study aims were developed and finalized jointly by the 3 authors (SH, RS, and GY) following background literature searches and a pilot study. Search terms and database specific−medical subject heading terms were used in various combinations (Multimedia Appendix 1). Boolean operators "OR" and "AND" were used to combine search terms.

Eligibility criteria were guided by the population, intervention, comparator, outcome, and study design framework [13] (Textbox 1). Titles and abstracts were reviewed for eligibility by 2 reviewers (SH and RS). Full-text papers were obtained and independently screened against the eligibility criteria by the same reviewers (SH and RS), and any disagreements were resolved through discussion. A third reviewer (GY) was available to resolve disagreements; however, this was not required. Manual searching of the reference lists was undertaken to identify any additional studies. The PRISMA flowchart details the search strategy for this review (Figure 1).

Textbox 1. Eligibility criteria.

- Inclusion criteria
    - Population
        - Adults (older than 18 years) with musculoskeletal conditions (acute, subacute, and chronic)
        - Setting: Anywhere patient has access to the internet
    - Intervention
        - Any form of digital-based intervention/treatment delivered by any digital means (eg, website or app) over any time frame.
        - Digital health intervention: For the purpose of this review, *digital health interventions* refers to *interventions for clients*, including targeted client communication; personal health tracking; and on-demand information services delivered by apps, web-based software, or websites. It includes any intervention accessed through a computer (work or home), or smartphone, or other hand-held device, and it includes desktop computer programs or apps that provide self-management information and can be used online or offline. The intervention must function without the need for directive input from a health professional. They must also be *interactive*, which we define as requiring contributions from program users (eg, entering personal data and making choices), which alter pathways within programs to produce tailored material and feedback that is personally relevant to users.
    - Comparator
        - The stated intervention(s) compared with waiting list control (no intervention) or alternative (*standard*) means of delivery (eg, face-to-face, class-based, and printed materials/hand-outs), nondigital self-management interventions (eg, leaflets), and noninteractive digital (eg, web page of flat copy).
    - Outcomes
        - Any positive or adverse health-based outcome and/or predictive indicators assessing pain and/or physical functioning/disability.
    - Secondary outcomes
        - Any positive or adverse health-based outcome and/or predictive indicators assessing patient knowledge and understanding, self-efficacy, catastrophizing, and empowerment. In addition, assess for any correlation between specific aspects of digital interventions and specific outcomes.
    - Study design
        - Randomized controlled trials in English.
- Exclusion criteria
    - Population
        - Nonmusculoskeletal pathology; postsurgical management, for example, following anterior cruciate ligament repair; and post knee replacement
        - Papers pertaining to (chronic) pain, where it is not possible to extrapolate information specifically relating to (chronic) pain of musculoskeletal origin
        - Papers that examined musculoskeletal pain as a *result* of computer use
    - Intervention
        - Papers where digital interventions are used in combination with other methods of intervention of nondigital origin and it is not possible to extrapolate the information pertaining specifically to digital interventions
    - Comparator
        - Any form of digital-based intervention/treatment
    - Outcome
        - Do not assess pain and/or physical functioning/disability
    - Study design
        - Study protocols, case studies/discussion papers, nonrandomized control trials, pilot studies, conference abstracts, and non-English language

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram indicating search strategy. CCT: controlled clinical trial; RCT: randomized controlled trial.



Data Extraction

Data relating to the research aims were independently extracted by the authors (SH,GY, RS). The data extracted included the study setting details (authors, year, and country of origin), the study population (number of participants, age, and gender), intervention details, duration/follow-up period, and outcome measures used (Table 1 and Multimedia Appendix 2). Any misunderstandings and disagreements were addressed through consultation.

All outcome measures and predictive tools (Multimedia Appendix 6) were acknowledged. There were numerous clinical outcome measures, with considerable disparity across the time frames over which interventions were assessed. In addition, a wide range of digital health interventions were used in the studies. According to Cochrane, this diversity across interventions and comparators is not compatible with statistical assessment via meta-analysis; instead, it is more suitable for narrative interpretation [13]. Attempting a meta-analysis with clinically diverse studies risks obscuring genuine differences in effects, resulting in inappropriate conclusions [13]. Furthermore, undertaking a meta-analysis of studies that are at risk of bias may be misleading, as this will compound the errors and produce results that may be interpreted inappropriately as having more credibility [13]. Therefore, for this review, a meta-analysis was not undertaken.

The results are presented in Multimedia Appendix 4. In line with the review’s primary aim, full details of study results are included in the table for all pain and functional disability outcomes. For additional outcomes, only significant between-group differences that were measured at the final time point in the study, are presented.

Table 1. Study and participant characteristics.

Reference	Subjects total, N	Subjects, n			Age (years), mean (SD)			Male: female (%)		
		CG[a]	EG[b]		CG	EG		CG	EG	
			EG1	EG2		EG1	EG2		EG1	EG2
Allen et al [14]	350	68	140[c]	142[d]	64.3 (12.2)	65.7 (10.3)[c]	65.3(11.5)[d]	22:78	29:71[c]	31:69[d]
Bennell et al [6]	148	74	74	__[e]	61.5 (7.6)	60.8 (6.5)	—	46:54	42:58	—
Bennell et al [15]	144	71	73	—	61.3 (7.1)	61.2 (7.2)	—	48:52	38:62	—
Bossen et al [16]	199	99	100	—	63 (5.4)	61 (5.9)	—	30:70	40:60	—
Buhrman et al [17]	54	28	26	—	42.9 (10.1)	43.5 (9.8)	—	36:64	27:73	—
Calner et al [18] and Nordin et al [19]	99	44	55	—	42 (11)	44 (10)	—	16:84	14:86	—
Carpenter et al [20]	141	71	70	—	42.5 (10.3)[f]	—	—	17:83[f]	—	—
Chhabra et al [21]	93	48	45	—	41.0 (14.2)	41.4 (14.2)	—	__[e]	—	—
Chiauzzi et al [22]	199	104	95	—	45.05 (11.72)	47.34 (12.23)	—	32:68	33:67	—
Del Pozo-Cruz et al [23-25]	100	50	50	—	45.5 (7.02)	46.83 (9.13)	—	11: 89	15:85	—
Irvine et al [26]	597	199	199[g]	199[h]	—	—	—	37:63	42:58[g]	41:59[h]
Krein et al [27]	229	118	111	—	51.9 (12.8)	51.2 (12.5)	—	86:14	89:11	—
Marangoni [28]	68	23	22[i]	23[j]	—	—	—	—		—
Mecklenberg et al [29]	125	54	101	—	47 (12)	46 (12)	—	74:26	57:43	—
Peters et al [30]		50	112[k]	114[l]	50.6 (10.1)	48.7 (11.5)[i]	47.5 (13.2)[j]	12:88	15:85[i]	17:83[j]
Petrozzi et al [31]	276	54	54	—	50.6 (14.4)	50.1 (12.8)	—	41:59	46:54	—
Shebib et al [32]	177	64	113	—	43 (12)	43 (11)	—	52:48	63:37	—
Toelle et al [33]	94	46	48	—	43 (11.0)	41 (10.6)	—	33:67	27:73	—
Van den Heuvel et al [34]	268	90	97	81	—	—	—	—	—	—

[a] CG: control group.

[b] EG: experimental group.

[c] PT: physical therapy.

[d] IBET: internet-based exercise training.

[e] not recorded.

[f] Total sample. Not recorded for control group and evaluation group.

[g] Fit back.

[h] Alternative care.

[i] CASP: computer-assisted stretching program.

[j] FLIP: facsimile lesson with instructional pictures (hard copy).

[k] iCBT: internet-delivered cognitive behavioral therapy.

[l] PPI: positive psychology intervention.

Quality Assessment

Methodological quality of the included studies was assessed by the lead researcher (SH) using the Cochrane risk of bias tool (modified) for quality assessment of randomized controlled trials (RCTs) [35]. This tool examines different subsets of bias, including performance, selection, detection, and attrition [36]. In total, 40% of the papers were independently assessed by a second reviewer (RS), and any disagreements were resolved through discussion. A third author was available (GY) should disagreements not be resolved but was not required.

Results

Search Results

A total of 1047 papers of potential interest were identified (Figure 1 and Multimedia Appendix 3). Of these, 301 were excluded as duplicates, leaving 746 for title and abstract screening. Following screening, the full texts of 57 papers were obtained and screened for eligibility, resulting in 22 papers eligible for inclusion.

Description of the Included Studies

Of the 22 papers identified (Figure 1), 3 papers were published in 1 study [23-25], whereas another study published 2 papers [18,19]. Although the study population and intervention were the same in Del Pozo et al's [23-25], Calner et al's, and Nordin et al's [18,19] papers, different outcomes were reported in each publication. Therefore, for the quality assessment, each of these papers was assessed individually using the designated assessment tool. However, for data extraction, the 3 Del Pozo et al's papers [23-25], and the Calner et al and Nordin et al's [18,19] papers, have been combined to avoid duplication of the participant outcomes, thus leaving 19 individual studies included in the data extraction and results tables (Table 1, and Multimedia Appendices 2, and 4). Of the 19 included studies, most were from the United States (n=8) [14,20,22,26-29,32]; 3 were from Australia [6,15,31]; 3 were from the Netherlands [16,30,34]; 2 were from Sweden [17-19]; and 1 each from Spain [23-25], Germany [33], and India [21]. One study was published in 2003 [34]; otherwise, all other studies were published between 2010 and 2019.

Of the studies that reported on the gender of participants, all studies except 3 [27,29,32] had a greater number of female participants. All studies included participants with an average age of 35 to 69 years. A wide variety of musculoskeletal conditions were examined by the included studies. In total, 10 studies investigated digital health interventions for LBP [17,20-27,31-33], 3 studies examined musculoskeletal pain [18,19,28,30], 3 studies investigated knee pain [6,14,29], 2 studies examined hip pain only or both knee and hip pain [15,16], and 1 investigated neck/upper limb pain [34].

Quality Assessment and Risk of Bias

Risk of bias was assessed using the modified Cochrane collaboration tool for assessing risk of bias [35]. Bias is assessed as a judgment (high, low, or unclear) for individual elements (7 domains) from 5 categories: selection (allocation concealment and randomization procedure), blinding (participants/personal and outcome assessors), completeness of data, selective outcome reporting, and other potential sources of bias [35]. Overall, the methodological quality of the included studies was variable (Table 2; refer to the judging criteria described in Multimedia Appendix 5). Of the 22 included studies, 1 achieved a low risk of bias across all 7 domains [15]. A further 6 were unable to blind the participants [6,14,21,23-25,27,31] but achieved a low risk of bias over the remaining 6 domains. Of the remaining 13 studies, a low risk of bias was achieved in 5 or fewer domains. The most consistent domain that failed to achieve a low risk of bias was the blinding of the study participants. The only study that was able to achieve blinding of participants [15] is one where both the control and the experimental group were told they would receive web-based resources and physiotherapy. However, 1 of the groups received only 8 information sheets on the web (flat copies; see Textbox 1) as opposed to the experimental group, which received an interactive digital intervention—pain coping skills training.

Table 2. Modified Cochrane collaboration tool for assessing risk of bias (For all domains, if reported Yes, this would indicate a low risk of bias, No would indicate a high risk of bias, and Unclear would indicate an unclear risk of bias).

Reference	Sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessors	Incomplete outcome data	Selective outcome reporting	Other sources of bias
Allen et al [14]	Yes	Yes	No	Yes	Yes	Yes	Yes
Bennell et al [6]	Yes	Yes	No	Yes	Yes	Yes	Yes
Bennell et al [15]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bossen et al [16]	Unclear	Yes	No	No	Yes	No	No
Buhrman et al [17]	Yes	Yes	No	Unclear	Yes	Yes	Yes
Calner et al [18]	Yes	Yes	No	Unclear	Yes	Yes	No
Carpenter et al [20]	Yes	Unclear	Unclear	Unclear	Yes	Yes	No
Chhabra et al [21]	Yes	Yes	No	Yes	Yes	Yes	Yes
Chiauzzi et al [22]	Yes	Yes	No	No	Yes	Yes	No
Del Pozo-Cruz et al [23]	Yes	Yes	No	Yes	Yes	Yes	Yes
Del Pozo-Cruz et al [24]	Yes	Yes	No	Yes	Yes	Yes	Yes
Del Pozo-Cruz et al [25]	Yes	Yes	No	Yes	Yes	Yes	Yes
Irvine et al [26]	Unclear	Unclear	No	No	Yes	Yes	No
Krein et al [27]	Yes	Yes	No	Yes	Yes	Yes	Yes
Marangoni [28]	No	No	Unclear	No	Unclear	Yes	No
Mecklenberg et al [29]	Yes	Yes	No	No	Yes	Yes	Yes
Nordin et al [19]	Yes	Yes	No	Unclear	Yes	Yes	No
Peters et al [30]	Unclear	Unclear	Unclear	Unclear	Yes	Yes	Unclear
Petrozzi et al [31]	Yes	Yes	No	Yes	Yes	Yes	Yes
Shebib et al [32]	Yes	Yes	No	No	Yes	Yes	No
Toelle et al [33]	No	No	Yes	No	Yes	Yes	Yes
Van de Heuvel et al [34]	Yes	Unclear	No	Unclear	Unclear	Yes	No

Pain

All 19 included studies used outcomes that assessed pain. Of these, 9 reported statistically significant improvements in pain [6,15,16,23-26,28,29,32,33], 4 studies [6,15,16,26] assessed pain at more than one time point, and in 2 studies [15,16], improvement was not maintained at the last time point measured. In all, 4 studies [15,16,23-26] reported effect sizes, measured by odds ratio, eta squared, or Cohen d (Multimedia Appendix 4). The findings in 1 study indicated a large effect size [23-25], 2 studies reported moderate effect sizes [15,26], and 1 study reported a small effect size [16]. In relation to the quality of the studies, 3 out of the 9 studies with positive outcomes [6,15,23-25] were within those classified as low risk of bias or those that had low risk of bias in 6 out of 7 categories, with the exception of blinding [6,23-25] (Table 2).

Of the 9 studies that demonstrated statistically significant improvement in pain, 1 was on participants with hip OA, 1 was on chronic knee and hip OA, 1 was on knee pain, and 5 were on LBP. Of the 10 studies that did not demonstrate improvement, 1 was on OA knee, 6 were on chronic LBP, 2 were on chronic musculoskeletal pain, and 1 was on work-related neck and upper limb disorders. The duration of the interventions varied from 3 weeks to 9 months in the studies that showed improvement and 3 weeks to 12 months in those studies that did not show positive outcomes.

Functional Disability

In all, 16 of the included studies [6,14-16,18-27,29-33] used outcomes that assessed functional disability. Of these, 10 reported statistically significant improvements [6,15,16,20,21,23-27,29,32] (Multimedia Appendix 4); 5 of these studies [6,15,16,26,27] assessed outcomes at more than one time point, and in 2 studies [15,27], improvement was not maintained at the last time point measured. A total of 6 studies [15,16,20,21,23-26] reported effect sizes. In total, 3 studies reported a large effect size [15,16,23-25], 2 studies reported a moderate effect size [20,21], and 1 study reported a small effect size [26] (Multimedia Appendix 4). In relation to the quality of the studies, 5 out of the 10 studies with positive outcomes [6,15,21,23-25,27] had a low or relatively low risk of bias (Table 2).

Of the 10 studies that demonstrated statistically significant improvement in functional disability, 2 studies were on participants with chronic knee pain, 1 was on hip OA, 1 was on chronic knee and hip OA, 3 were on LBP, and 3 were on chronic LBP. Of the 6 studies that did not demonstrate improvement, 1 was on OA knee, 1 was on LBP, 2 were on chronic LBP, and 2 were on chronic musculoskeletal pain [14,18,19,22,30,31,33]. The duration of the interventions varied from 3 weeks to 9 months in the studies that showed improvement and 4 weeks to 4 months in those studies that did not show positive outcomes.

Additional Outcome Measures

Several additional outcomes were measured across the studies including measures of quality of life, psychosocial distress, work, and surgery interest. Multimedia Appendix 2 gives details of all outcomes, and Multimedia Appendix 4 shows statistically significant results at the last measured time point for each study. The most frequent additional outcomes across the studies were catastrophizing, self-efficacy, quality of life, and coping strategies.

Of the 7 studies reporting on catastrophizing [6,15,17,20,22,31,33], 4 reported statistically significant improvements [6,15,17,20]. Moreover, 7 studies examined self-efficacy [6,15,16,18-20,22,31], 2 of which reported statistically significant improvements [6,20]. In all, 6 studies reported on health-related quality of life [6,15,17-19,23-26], of which 3 [6,17,23-25] reported statistically significant improvements. In total, 6 studies reported on coping ability [6,15,16,18,19,22,30], of which 4 [6,15,16,22] showed significant improvement. A total of 2 studies examined self-reported interest in surgery post intervention [29,32], and both the studies showed significant reductions in pursuing surgical intervention.

Discussion

Principal Findings

The primary aim of this systematic review was to assess if digital health interventions were clinically effective in impacting musculoskeletal pain and functional disability in patients with musculoskeletal conditions. The results of the analysis show that there is some evidence to support the effectiveness of digital health interventions in improving pain, with 9 out of 19 studies reporting significant improvements. There was stronger evidence to support the role of digital health interventions in improving functional disability, with 10 out of 16 studies reporting significant improvements. There were also positive results shown in several additional outcomes, most notably catastrophizing and coping strategies, with 4 out of 7 and 4 out of 6 studies, respectively, reporting significant improvements.

In terms of musculoskeletal conditions treated, both peripheral and spinal conditions showed improvement in pain and/or functional disability. However, pain outcomes in all studies with a study population of chronic musculoskeletal conditions (chronic low back and chronic musculoskeletal pain) did not show any significant improvements. This is not a surprising finding, as generally interventions for chronic low back and chronic musculoskeletal pain are likely to have a greater effect on function, quality of life, and psychosocial factors rather than pain [37], as was found in this review. The positive outcomes for the majority of studies that considered coping strategies and catastrophizing may indicate that digital education and management strategies enable patients to better understand and cope with their musculoskeletal condition. The reduction in interest in surgery found in 2 studies [29,32] supports this.

Overall, the methodological quality of the included trials was variable. Only 1 study [15] had a low risk of bias across all domains. In total, 6 further studies were unable to blind the study participants [6,14,21,23-25,27,31], which is a common challenge facing researchers developing pragmatic clinical trials with comparative interventions [36]. Therefore, the potential risk of performance bias is elevated [36]. However, there was no observed direct relationship between the quality of trials and positive or negative outcomes; 3 of the 7 studies with low or relatively low risk of bias demonstrated significant improvements in pain, and 5 studies demonstrated significant improvements in functional disability. The large number of outcome measures used by the included studies made direct comparisons between studies difficult. This is a common problem in musculoskeletal research and is a reflection of the large number of outcome measures used in musculoskeletal conditions. The digital health interventions also varied considerably in many aspects including duration, program features, and targeted musculoskeletal condition, meaning a meta-analysis was not possible.

The secondary aim of the study was to explore the content, characteristics, and delivery of digital health interventions to ascertain if there are specific aspects of the interventions, or the population they are targeting, that are associated with beneficial outcomes. In relation to this aim, we looked across all studies in an attempt to identify characteristics related to positive or negative outcomes. Several features emerged following the analysis of the studies.

It appears important to match the digital health intervention to known evidence-based approaches for the condition. Examples of this can be seen in the studies where the population had chronic LBP or chronic musculoskeletal pain. In 3 of the studies that did not show any significant improvements in functional disability [17,22,30], interventions were not matched to what would be considered the best evidence-based practice. It is widely recognized that chronic musculoskeletal conditions, particularly LBP, are optimally managed using a biopsychosocial approach [38] incorporating both physical and psychosocial elements in the rehabilitation program. Chiauzzi et al [22], Peters et al [30], and Buhrman et al [17] utilized components of a psychosocial approach within their digital health interventions, but they did not specify an exercise or physical activity component. None of these studies demonstrated improvement in pain or functional disability. In contrast, 3 studies [20,21,26] did achieve statistically significant improvements in functional disability when including an exercise/physical activity component alongside a psychosocial component within their digital health intervention. In total, 2 studies [18,19,31] demonstrated the inclusion of all components of a biopsychosocial approach but achieved no statistically significant improvement in functional disability; however, this can be explained to an extent by the nature of the studies. In Calner et al's and Nordin et al's studies [18,19], the control group received multimodal rehabilitation (MMR) treatment from a minimum of 3 different health care professionals, including physiotherapists, psychologists, physicians, occupational therapists, and nurses. The experimental group also received MMR plus a web-based behavior change program; therefore, both the experimental and control groups had access to extensive psychosocial and physical intervention. Petrozzi et al [31] conducted an established internet-delivered program designed for the prevention and management of depressive symptoms (MoodGYM) and conducted a single-blinded study to examine the effectiveness of this in combination with physical treatments for patients with chronic LBP. The lack of significant improvement may be a reflection of the mismatch of content to the target population. The population for this study had moderate levels of back pain, low levels of disability, high levels of self-efficacy, and normal to mild levels of psychological distress (as assessed by STarTBack screening tool). However, MoodGYM is targeted toward those with higher levels of psychological distress and at higher risk of ongoing disability. The authors themselves acknowledge this as a limitation of their study. This highlights the importance of content being appropriately targeted toward the intended audience.

Another feature we identified was that all the digital health interventions delivered on an app [21,26,29,32,33], as opposed to a web-based program, produced positive results in pain and/or functional disability. In the context of this review, apps appear to have gained popularity in recent years. Of the 5 studies using apps, 1 was published in 2015 and the remaining 4 in 2018-2019. A number of reasons can be hypothesized as to why apps may provide successful digital health interventions. All the apps had in-app functions that facilitated greater engagement with study participants, for example, sensor-guided exercise features, notifications, and daily activity goals. Additionally, the success of apps may be related to other factors, including ease of access, portability, and convenience in comparison to web-based interventions.

In several studies, additional efforts were made to encourage engagement with the digital intervention. Various forms of multimedia additional support were included in 10 studies, such as phone calls, email reminders, and text messages [6,14-17,23-26,29-31]. In total, 6 of these studies demonstrated positive results for either pain and/or functional disability. Therefore, there is some indication that these additional forms of support may be linked to positive outcomes; however, the frequency and delivery modes were variable; as such, it is difficult to quantify the extent to which additional forms of support improve the effectiveness of a digital health intervention.

Due to the different features within each intervention, it was difficult to draw any firm conclusions regarding which components of digital intervention create the most engaging digital interface. The number of participant-interactive components within web-based interventions (eg, exercise trackers, web-based coaching, and quizzes) did not appear to definitively influence the success of the intervention. Both significant and nonsignificant outcomes were seen in trials with multiple interactive elements (Multimedia Appendix 2). An RCT by Riva et al [39] was designed to specifically evaluate the addition of interactive features to a well-established internet intervention for chronic back pain; the results of this study showed no difference between the group with multiple interactive features and the control in relation to pain and physical activity. It is also unclear if tailoring the intervention offers additional benefits. One of the studies included in this review [26] was a 3-arm study comparing a control group, tailored mobile-web intervention, and an alternative care group that received emails directing participants to nontailored web-based resources. Significant reductions in function were reported at 16-week follow-up for both intervention groups compared with the control; however, there was no significant difference between the groups.

In all studies reviewed, there was minimal reference to patient involvement in the development of the digital health intervention. Many studies on such interventions appear to use content developed by the medical/research team, with little reference to patient involvement in the design phase of the intervention [40]. Involving patients early in the development process may help inform key features of the design, including what constitutes an engaging interface.

In 6 studies [6,15,18,19,29,31,33], digital health interventions were used as an adjunct to face-to-face intervention with a health care practitioner, that is, physiotherapist or multidisciplinary team; 4 of these studies [6,15,29,33] showed improvements in pain and/or functional disability. In the 2 studies in which digital intervention did not show any improvement [18,19,31], the studies by Calner et al and Nordin et al [18,19] involved extensive MDT rehabilitation in both groups, and the addition of web-based intervention did not improve outcomes. This may reflect the intensity of face-to-face treatment received by the study participants in both groups. Petrozzi et al's study [31] was targeted at patients with high levels of psychosocial distress that did not match the presentation of the patients in the study group; the mismatch of the intervention to the target group may have influenced the results. As such, there is some support that digital health interventions may improve outcomes as an adjunct to face-to-face treatment with a health care practitioner. Importantly, in no cases were digital health interventions inferior to an interventional control in relation to pain or functional disability, and no trials reported adverse events. This, in itself, is an important finding, as although not always superior to interventional controls (usual care), digital health interventions have the ability to deliver safe, high-reach, low-cost, readily accessible, and scalable care. They could also help address physical access issues, as a result of the nature of a patient's pain, comorbidities, travel distances, and costs. Therefore, the use of digital interventions as an alternative to usual care may have a substantial impact on helping to manage the growing burden of musculoskeletal functional disability. This is particularly pertinent in health care systems currently stretched to such an extent that the frequency of delivery of face-to-face appointments is suboptimal. Digital health interventions may ultimately result in patients accessing more health care than they would in a solely face-to-face scenario.

From the review, no conclusions can be drawn regarding the duration of intervention for both pain and functional disability. Significant and nonsignificant outcomes were found in both short duration (3 weeks) and longer duration studies (up to 12 months). Due to the heterogeneity of the interventions and the lack of detail in the studies, it is also difficult to draw conclusions on the optimal dose or exposure to digital health interventions required to gain meaningful benefits. Many studies did not quantify how long patients engaged with the digital intervention; therefore, it is not possible to conclude if patients who engaged for longer durations with the intervention did better or worse. Further research is needed in both areas.

In what is perhaps a reflection of the emerging role of digital health interventions in the management of musculoskeletal conditions, there was a lack of long-term follow-up, particularly for patients with chronic musculoskeletal conditions in the majority of studies. It would be beneficial to assess the impact of successful websites and apps over a longer duration, as this has potential implications for patients, services, and health care resources if acute exacerbations of chronic conditions can be, at least in part, managed remotely.

There are certain limitations to this systematic review. Only English language studies were included; therefore, it is possible that relevant literature published in other languages may have been excluded. In addition, this review only included RCT study designs; however, this was to ensure that higher levels of evidence were used to address the aim of the review [41]. Finally, it was not possible to undertake a meta-analysis due to the diversity across interventions and comparators in the reviewed trials [13].

Conclusions

This review has demonstrated that digital health interventions have some clinical benefits in the management of musculoskeletal conditions. There is evidence to support the effectiveness of digital health interventions in improving pain. There is stronger evidence to support the effectiveness of digital health interventions in improving functional disability. There are also positive results shown in several additional outcomes, notably catastrophizing and coping strategies. This review demonstrates the potential of digital health interventions to contribute positively toward diminishing the personal, societal, and economic impact of musculoskeletal conditions, which, as our population ages, is only set to grow. Further research is needed to identify the patient subgroups that respond most positively to digital health interventions and also to determine the pertinent features of the interventions that are likely to achieve more successful patient outcomes.

Acknowledgments

The authors would like to thank their librarian, Suzanne Ford, at St Helens and Knowsley NHS Trust for supporting them with accurate literature search.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[DOCX File , 13 KB - jmir_v22i6e15617_app1.docx]

Multimedia Appendix 2
Intervention characteristics.
[DOCX File , 45 KB - jmir_v22i6e15617_app2.docx]

Multimedia Appendix 3
Search strategies.
[DOCX File , 23 KB - jmir_v22i6e15617_app3.docx]

Multimedia Appendix 4
Results for pain and functional disability outcomes and for significant additional outcomes.
[DOCX File , 31 KB - jmir_v22i6e15617_app4.docx]

Multimedia Appendix 5
Risk of bias assessment.
[DOCX File , 16 KB - jmir_v22i6e15617_app5.docx]

Multimedia Appendix 6

Abbreviations for outcomes measures.
[DOCX File , 17 KB - jmir_v22i6e15617_app6.docx]

References

1. Hill JC, Kang S, Benedetto E, Myers H, Blackburn S, Smith S, et al. Development and initial cohort validation of the Arthritis research UK musculoskeletal health questionnaire (MSK-HQ) for use across musculoskeletal care pathways. BMJ Open 2016 Aug 5;6(8):e012331 [FREE Full text] [doi: 10.1136/bmjopen-2016-012331] [Medline: 27496243]
2. World Health Organization. 2019. Musculoskeletal Conditions URL: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions [accessed 2020-01-26]
3. Versus Arthritis. 2019. The State of Musculoskeletal Health 2019: Arthritis and Other Musculoskeletal Conditions in Numbers URL: https://www.versusarthritis.org/media/14594/state-of-musculoskeletal-health-2019.pdf [accessed 2020-01-26]
4. Skills for Health. 2018. Musculoskeletal Core Capabilities Framework URL: https://www.skillsforhealth.org.uk/services/item/574-musculoskeletal-core-skills-framework [accessed 2020-04-09]
5. Bone and Joint Initiative USA. 2016. The Impact of Musculoskeletal Disorders on Americans-Opportunities for Action URL: http://www.boneandjointburden.org/docs/BMUSExecutiveSummary2016.pdf [accessed 2020-03-06]
6. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
7. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011.
8. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019.
9. The Lancet. Is digital medicine different? Lancet 2018 Jul 14;392(10142):95. [doi: 10.1016/S0140-6736(18)31562-9] [Medline: 30017135]
10. Greaves F, Joshi I, Campbell M, Roberts S, Patel N, Powell J. What is an appropriate level of evidence for a digital health intervention? Lancet 2019 Dec 22;392(10165):2665-2667. [doi: 10.1016/S0140-6736(18)33129-5] [Medline: 30545779]
11. The Government of UK. 2018. The Future of Healthcare: Our Vision for Digital, Data and Technology in Health and Care URL: https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care [accessed 2020-01-26]
12. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
13. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. The Cochrane Collaboration. 2011. URL: http://handbook-5-1.cochrane.org [accessed 2020-01-26]
14. Allen KD, Arbeeva L, Callahan LF, Golightly YM, Goode AP, Heiderscheit BC, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis Cartilage 2018 Mar;26(3):383-396 [FREE Full text] [doi: 10.1016/j.joca.2017.12.008] [Medline: 29307722]
15. Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S, et al. Effects of internet-based pain coping skills training before home exercise for individuals with hip osteoarthritis (HOPE trial): a randomised controlled trial. Pain 2018 Sep;159(9):1833-1842. [doi: 10.1097/j.pain.0000000000001281] [Medline: 29794609]
16. Bossen D, Veenhof C, van Beek KE, Spreeuwenberg PM, Dekker J, de Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
17. Buhrman M, Nilsson-Ihrfeldt E, Jannert M, Ström L, Andersson G. Guided internet-based cognitive behavioural treatment for chronic back pain reduces pain catastrophizing: a randomized controlled trial. J Rehabil Med 2011 May;43(6):500-505 [FREE Full text] [doi: 10.2340/16501977-0805] [Medline: 21533329]
18. Calner T, Nordin C, Eriksson MK, Nyberg L, Gard G, Michaelson P. Effects of a self-guided, web-based activity programme for patients with persistent musculoskeletal pain in primary healthcare: a randomized controlled trial. Eur J Pain 2017 Jul;21(6):1110-1120. [doi: 10.1002/ejp.1012] [Medline: 28464364]
19. Nordin CA, Michaelson P, Gard G, Eriksson MK. Effects of the web behavior change program for activity and multimodal pain rehabilitation: randomized controlled trial. J Med Internet Res 2016 Oct 5;18(10):e265 [FREE Full text] [doi: 10.2196/jmir.5634] [Medline: 27707686]
20. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
21. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
22. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]

23. del Pozo-Cruz B, Parraca JA, del Pozo-Cruz J, Adsuar JC, Hill J, Gusi N. An occupational, internet-based intervention to prevent chronicity in subacute lower back pain: a randomised controlled trial. J Rehabil Med 2012 Jun;44(7):581-587 [FREE Full text] [doi: 10.2340/16501977-0988] [Medline: 22674240]
24. del Pozo-Cruz B, Gusi N, del Pozo-Cruz J, Adsuar JC, Hernandez-Mocholí M, Parraca JA. Clinical effects of a nine-month web-based intervention in subacute non-specific low back pain patients: a randomized controlled trial. Clin Rehabil 2013 Jan;27(1):28-39. [doi: 10.1177/0269215512444632] [Medline: 22653374]
25. del Pozo-Cruz B, Adsuar JC, Parraca J, del Pozo-Cruz J, Moreno A, Gusi N. A web-based intervention to improve and prevent low back pain among office workers: a randomized controlled trial. J Orthop Sports Phys Ther 2012 Oct;42(10):831-841. [doi: 10.2519/jospt.2012.3980] [Medline: 22951407]
26. Irvine AB, Russell H, Manocchia M, Mino DE, Glassen TC, Morgan R, et al. Mobile-web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 2;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
27. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013 Aug 19;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
28. Marangoni AH. Effects of intermittent stretching exercises at work on musculoskeletal pain associated with the use of a personal computer and the influence of media on outcomes. Work 2010;36(1):27-37. [doi: 10.3233/WOR-2010-1004] [Medline: 20555173]
29. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
30. Peters ML, Smeets E, Feijge M, van Breukelen G, Andersson G, Buhrman M, et al. Happy despite pain: a randomized controlled trial of an 8-week internet-delivered positive psychology intervention for enhancing well-being in patients with chronic pain. Clin J Pain 2017 Nov;33(11):962-975 [FREE Full text] [doi: 10.1097/AJP.0000000000000494] [Medline: 28379873]
31. Petrozzi MJ, Leaver A, Ferreira PH, Rubinstein SM, Jones MK, Mackey MG. Addition of MoodGYM to physical treatments for chronic low back pain: a randomized controlled trial. Chiropr Man Therap 2019;27:54 [FREE Full text] [doi: 10.1186/s12998-019-0277-4] [Medline: 31673330]
32. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
33. Toelle T, Utpadel-Fischler D, Haas K, Priebe J. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
34. van den Heuvel SG, de Looze MP, Hildebrandt VH, Thé KH. Effects of software programs stimulating regular breaks and exercises on work-related neck and upper-limb disorders. Scand J Work Environ Health 2003 Apr;29(2):106-116 [FREE Full text] [doi: 10.5271/sjweh.712] [Medline: 12718496]
35. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
36. Kamper SJ. Bias: linking evidence with practice. J Orthop Sports Phys Ther 2018 Aug;48(8):667-668. [doi: 10.2519/jospt.2018.0703] [Medline: 30064332]
37. Guidelines for Pain Management Programmes for Adults. Central London, UK: The British Pain Society; 2013.
38. Kamper S, Apeldoorn A, Chiarotto A, Smeets R, Ostelo R, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2014 Sep 2(9):CD000963. [doi: 10.1002/14651858.CD000963.pub3] [Medline: 25180773]
39. Riva S, Camerini A, Allam A, Schulz PJ. Interactive sections of an Internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014 Aug 13;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
40. Atack L, Luke R. The impact of validated, online health education resources on patient and community members' satisfaction and health behaviour. Health Educ J 2011 Jan 7;71(2):211-218. [doi: 10.1177/0017896910393785]
41. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 2011 Jul;128(1):305-310 [FREE Full text] [doi: 10.1097/PRS.0b013e318219c171] [Medline: 21701348]

Abbreviations

LBP: low back pain
MMR: multimodal rehabilitation
OA: osteoarthritis

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 26.07.19; peer-reviewed by N Wilson, A Kora, I Castrejon, SGS Shah, E Sadeghi-Demneh; comments to author 21.10.19; revised version received 02.02.20; accepted 22.03.20; published 05.06.20.

Please cite as:
Hewitt S, Sephton R, Yeowell G
The Effectiveness of Digital Health Interventions in the Management of Musculoskeletal Conditions: Systematic Literature Review
J Med Internet Res 2020;22(6):e15617
URL: https://www.jmir.org/2020/6/e15617
doi:10.2196/15617
PMID:32501277

Original Paper

Efficacy of Three Low-Intensity, Internet-Based Psychological Interventions for the Treatment of Depression in Primary Care: Randomized Controlled Trial

Margalida Gili[1,2,3], PhD; Adoración Castro[1,2], PhD; Azucena García-Palacios[4,5], PhD; Javier Garcia-Campayo[3,6], PhD; Fermin Mayoral-Cleries[7], PhD; Cristina Botella[4,5], PhD; Miquel Roca[1,2,3], PhD; Alberto Barceló-Soler[8], PhD; María M Hurtado[7], PhD; MªTeresa Navarro[9], PhD; Amelia Villena[10], ClinPsy; M Ángeles Pérez-Ara[1,2], PhD; Pau Riera-Serra[1,2], MSc; Rosa Mª Baños[5,11], PhD

[1]Institut Universitari d'Investigació en Ciències de la Salut, University of Balearic Islands, Palma de Mallorca, Spain
[2]Institut d'Investigació Sanitaria Illes Balears, Palma de Mallorca, Spain
[3]Primary Care Prevention and Health Promotion Research Network, RedIAPP, Madrid, Spain
[4]Department of Clinical and Basic Psychology and Biopsychology, Faculty of Health Sciences, Univeristy Jaume I, Castellón, Spain
[5]Biomedical Research Center Network (CIBER) Physiopathology Obesity and Nutrition (CIBERobn), Carlos III Health Institute, Madrid, Spain
[6]Departament of Psychiatry, Hospital Miguel Servet, University of Zaragoza, Zaragoza, Spain
[7]Mental Heath Unit, Hospital Regional of Malaga, Biomedicine Research Institute (IBIMA), Málaga, Spain
[8]Aragón Institute for Health Research (IIS Aragon), Zaragoza, Spain
[9]Department of Psychology and Sociology, University of Zaragoza, Zaragoza, Spain
[10]Mental Health Unit of Pozoblaco, Hospital Los Pedroches, Córdoba, Spain
[11]Department of Psychological, Personality, Evaluation and Treatment, University of Valencia, Valencia, Spain

Corresponding Author:
Margalida Gili, PhD
Institut Universitari d'Investigació en Ciències de la Salut
University of Balearic Islands
Carretera de Valldemossa
Palma de Mallorca, 07122
Spain
Phone: 34 173081
Email: mgili@uib.es

Abstract

Background: Primary care is a major access point for the initial treatment of depression, but the management of these patients is far from optimal. The lack of time in primary care is one of the major difficulties for the delivery of evidence-based psychotherapy. During the last decade, research has focused on the development of brief psychotherapy and cost-effective internet-based interventions mostly based on cognitive behavioral therapy (CBT). Very little research has focused on alternative methods of treatment for depression using CBT. Thus, there is a need for research into other therapeutic approaches.

Objective: This study aimed to assess the effectiveness of 3 low-intensity, internet-based psychological interventions (healthy lifestyle psychoeducational program [HLP], focused program on positive affect promotion [PAPP], and brief intervention based on mindfulness [MP]) compared with a control condition (improved treatment as usual [iTAU]).

Methods: A multicenter, 4-arm, parallel randomized controlled trial was conducted between March 2015 and March 2016, with a follow-up of 12 months. In total, 221 adults with mild or moderate major depression were recruited in primary care settings from 3 Spanish regions. Patients were randomly distributed to iTAU (n=57), HLP (n=54), PAPP (n=56), and MP (n=54). All patients received iTAU from their general practitioners. The main outcome was the Spanish version of the Patient Health Questionnaire-9 (PHQ-9) from pretreatment (time 1) to posttreatment (time 2) and up to 6 (time 3) and 12 (time 4) months' follow-up. Secondary outcomes included the visual analog scale of the EuroQol, the Short-Form Health Survey (SF-12), the Positive and Negative Affect Schedule (PANAS), and the Pemberton Happiness Index (PHI). We conducted regression models to estimate outcome differences along study stages.

Results: A moderate decrease was detected in PHQ-9 scores from HLP (β=–3.05; *P*=.01) and MP (β=–3.00; *P*=.01) compared with iTAU at posttreatment. There were significant differences between all intervention groups and iTAU in physical SF-12 scores at 6 months after treatment. Regarding well-being, MP and PAPP reported better PHI results than iTAU at 6 months post treatment. PAPP intervention significantly decreased PANAS negative affect scores compared with iTAU 12 months after treatment.

Conclusions: The low-intensity, internet-based psychological interventions (HLP and MP) for the treatment of depression in primary care are more effective than iTAU at posttreatment. Moreover, all low-intensity psychological interventions are also effective in improving medium- and long-term quality of life. PAPP is effective for improving health-related quality of life, negative affect, and well-being in patients with depression. Nevertheless, it is important to examine possible reasons that could be implicated for PAPP not being effective in reducing depressive symptomatology; in addition, more research is still needed to assess the cost-effectiveness analysis of these interventions.

Trial Registration: ISRCTN Registry ISRCTN82388279; http://www.isrctn.com/ISRCTN82388279

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-015-0475-0

(J Med Internet Res 2020;22(6):e15845) doi:10.2196/15845

KEYWORDS

depression; primary care; internet-based interventions; randomized controlled trial

Introduction

Background

Depression represents a significant personal, economic, and societal burden [1-3]. Primary care remains a major access point for initial treatment of depression [4,5]. Previous studies reveal that more than 80% of patients with depression are being managed in general practice [6,7]. However, management of these patients is far from optimal, and it has been reported that only one-half of the patients receive adequate care, whether pharmacological or psychological [5]. Pence et al [4] estimated that only 47% of primary care patients with depression are clinically recognized, 24% receive treatment, and 9% receive adequate treatment.

Multiple and complex facilitators and barriers to treatment have been described [8,9], and access to evidence-based psychotherapy is one of these difficulties. Different factors such as professionals' training, time needed, costs or work overload, professionals' attitudes and organization, and geographical and logistic difficulties are some of the reasons that make the integration of psychotherapy in primary care difficult. New forms and models for delivering psychotherapy in primary care have been proposed to overcome these problems. The implementation of stepped and collaborative care models in primary care settings that provide time-limited psychotherapy have been found to improve the management of depression [10]. One of the most important difficulties for integrating psychotherapy into primary care is the lack of time and resources. Many empirically tested treatment protocols last 1 hour on a weekly basis for 15-20 sessions. (Extensive time and resources imply important difficulties for the application of this type of therapy.) For these reasons, and in an effort to reduce the high medical costs of depression treatment and overcome the difficulties of traditional treatments in primary care, brief psychotherapy for depression effective and cost-effective internet-based interventions have been extensively developed during the last decade [11,12]. Low-intensity, internet-based interventions could be a simple, cost-effective method for treating depression in primary care settings [13]. In fact, meta-analyses suggest that depression can be effectively treated with brief psychotherapy (6-8 sessions), specifically with cognitive behavioral therapy (CBT), problem-solving therapy [14], and counseling approaches [15].

Most of the internet interventions aiming at the treatment of depression are based on CBT. Previous findings for other forms of face-to face psychotherapy suggest that there is no *one-fits-all* solution [16], but few studies have analyzed web-based interventions based on other types of treatments. First results are promising, but more research is needed to determine the efficacy of alternative programs than internet-delivery CBT [17].

In a previous study, our group shows the efficacy of an internet intervention for depression in primary care (smiling is fun) [18]. The program follows a transdiagnostic perspective, and it is based on CBT techniques but also includes other psychological strategies to improve depressive symptoms such as promotion of healthy lifestyles, positive affect, and mindfulness. The treatment protocol is composed of 10 modules and lasts about 3 months. The program proved to be more effective than treatment as usual alone, but the attrition rate at follow-up was significant and the retention rate was not as good as we expected. These results led us to design a new protocol with the objective to identify which of the alternative therapeutic approaches was more effective and, also, to shorten the duration of the program to achieve better rates of attrition and retention.

Evidence of the benefits for treating depression of positive psychology, mindfulness, and lifestyle habits delivered using internet is growing as a result of increase in research studies over the last 10 years [19-22]. Meta-analyses of positive psychological interventions delivered to clinical and community samples have reported small but significant effects for reducing depressive symptoms [23] and comparable efficacy and lasting effects to traditional psychotherapy and pharmacotherapy [24,25]. Web-based positive emotion skills training for depression have shown preliminary good results [19,26] but warrant additional study. There is growing evidence for positive benefits of mindfulness interventions in clinically depressed individuals [27,28]. Recent reviews and meta-analysis of mindfulness-based CBT suggest that internet-based approaches have potential to contribute to improving depression [29,30], but there is a need for research into this therapeutic approach. Finally, lifestyle web-based interventions can be an effective and inexpensive alternative or supplement to depression therapy that is delivered using more traditional modes, overcoming barriers that make people from accessing treatment difficult. The results of a recent review [31] highlight the potential of web-based lifestyle interventions as adjunctive treatments for depression and the possibility of achieving significant improvements in depressive symptoms when targeting lifestyle behavior change. But the limited number of studies requires further clinical trials to achieve better understanding.

Objectives

Considering the scarcity of these studies and the fact that low-intensity, internet-based psychological interventions could be an efficacious and cost-effective therapeutic option for the treatment of depression, the aim of this study was to assess the effectiveness of 3 low-intensity, internet-based psychological interventions (psychoeducational program for the promotion of a healthy lifestyle (HLP), psychological intervention for the promotion of positive affect (PAPP), and brief intervention based on mindfulness [MP]) compared with a control condition.

Methods

Study Design

This study was a multicenter, 4-arm, parallel randomized controlled trial. Adults with depressive symptoms in primary care were randomly assigned to one of the following groups: (1) HLP + improved treatment as usual (iTAU), (2) PAPP + iTAU, (3) MP + iTAU, or (4) iTAU.

Trial registration number of this study was ISRCTN82388279. Research protocol of the study has been described elsewhere [32].

Recruitment of Participants and Baseline Assessment

We recruited patients with major depression or dysthymia, older than 18 years, able to understand and read Spanish, with mild or moderate depression according to the Patient Health Questionnaire-9 (PHQ-9; 5-9: mild depression; 10-14: moderate depression) [33], and with symptoms lasting longer than 2 weeks. Major depression and dysthymia were identified using the MINI International Neuropsychiatric Interview 5.0. We excluded patients with a diagnosis of any disease that may affect the central nervous system (brain pathology, traumatic brain injury, dementia, etc); with any psychiatric disorder other than major depression, dysthymia, anxiety disorders, or personality disorders; with any medical, infectious, or degenerative disease that may affect mood; with presence of delusional ideas or hallucinations consistent or not with mood; and with suicide risk.

Participants were recruited in primary care settings, between March 2015 and March 2016, in the Spanish regions of Aragon, Andalusia, and the Balearic Islands. When the general practitioner identifies a potential participant during a routine visit, he or she explained to the patient the characteristics of the study. When the patient was interested in participating, he or she signed an informed consent form and the general practitioner filled a referral form describing the sociodemographic characteristics of the patient and a checklist for inclusion and exclusion criteria and gave him or her the patient's information sheet and a handout describing the study. The general practitioner sent these documents by fax to the local researcher. Participants were interviewed in the next 3 days by the researcher, which administered psychological assessment instruments related with inclusion and exclusion criteria by phone. Included participants were randomized to 1 of the 4 groups by an independent researcher. Patient safety was systematically monitored. The Ethical Review Board of the regional health authority approved the study (Ref: IB 2144/13PI).

Randomization, Concealment, and Blinding

The sequence was concealed until interventions were assigned. Participants agreed to participate before the random allocation without knowing which treatment they were being allocated to. Study personnel conducting psychological assessment were masked to participants' treatment conditions. The researcher that administered baseline assessments was unaware of the treatment group to which the participant belonged. This researcher was different from the one that administered the questionnaires over the study. General practitioners were also unaware, as far as possible, of the arm to which each patient had been randomized, as their treatment needed to be exclusively based on the recommendations of the treatment of depression guidelines.

Follow-Up

Follow-up data collection took place between April 2015 and June 2017. Participants were assessed on web at pretreatment (time 1), posttreatment (time 2), and 6- (time 3) and 12- (time 4) month posttreatment assessments. The web-based platform hosted the questionnaires. Participants were sent an email with a link to that platform. No other protocols were used to increase compliance with the research data collection, but a phone call was made before each wave assessment to increase response rates.

Improved Treatment as Usual

All the patients included in the study (irrespective of the treatment group randomly assigned) received iTAU. This treatment was provided by their general practitioners, who had previously received a training program to update their knowledge on how to diagnose and treat depression in primary care and optimized by the recommendations based on the Spanish Guide for the Treatment of Depression in Primary Care [34,35]. In case of suicide risk or severe social dysfunction or worsening of symptoms being detected, patients were referred to mental health facilities.

Intervention Groups

All interventions (except iTAU) were composed of one face-to-face group session and 4 web-based, individual, and interactive therapeutic modules.

The face-to-face session, which took place in primary care centers, involved up to 5 patients and was 90 min long. The aim of this session was to explain the program structure and main components of treatment and to motivate participants for change.

The web-based therapeutic modules are oriented to work on different psychological techniques, and the duration of each module is approximately between 40 and 60 min. All modules include an explanation of the module contents, check questions to test if they understand the contents, and exercises to practice the techniques. These modules are sequential, to move step by step, throughout the program. However, users can review the module contents once they are finished. Although the duration of the program can vary among users, it is estimated that for most people, it lasted between 4 and 8 weeks. Regarding the therapeutic content, all intervention groups are composed of 4 intervention modules based on different psychological techniques, as shown in Table 1. A more specific and detailed description of the module contents can be found elsewhere [32].

To maximize adherence, participants received 2 weekly automated mobile phone messages, encouraging them to proceed with the program and reminding them of the importance of doing the tasks in each module. If participants did not access the program for a week, they received an automated email encouraging them to continue with the modules. Furthermore, the program also offers continued feedback to users through the assessment tools showing them their progress throughout the entire treatment process. All groups of patients received a participant manual with information about the technical aspects of the web-based program.

Table 1. Intervention modules and main objectives.

Intervention and modules	Main objective
Psychoeducational program for the promotion of a healthy lifestyle	
Beginning of a lifestyle change	To teach the importance of healthy lifestyle to improve emotional health and general well-being and to give structured hygienic-dietary recommendations.
Physical activity. Learning to move on	To give information about the most recommended exercises to improve mood, and to train the patient in learning procedures to increase motivation, to start being more active, and to maintain this physical activity regularly.
Diet. Learning to eat	To teach the importance of diet to achieve a good physical and mental health, and the role of the Mediterranean diet in the prevention and treatment of depression.
Sleep. The importance of good sleep	To understand the relationship between sleep and general health.
Psychological intervention for the promotion of positive affect	
Learning to live	To teach the importance of establishing and maintaining an adequate activity level and the relevance of choosing activities that are significant, with a personal meaning for the individual.
Learning to enjoy	To give education about the effect of positive emotions and to train the patient in learning procedures to increase the likelihood of experiencing positive emotions, promoting the occurrence of pleasant activities to learn to enjoy the present moment.
Accepting to life	To train the patient in focusing on positive emotions related with the past (such as gratitude) or the future (such as optimism).
Living and learning	To train the patient in understanding life as a continuous process of learning and personal growth, emphasizing the training in strategies to promote psychological strengths, resilience, and meaningful goals linked to important values.
Brief intervention based on mindfulness	
Getting to know mindfulness	To show what mindfulness is, prejudices about it, the inattention problem, and some of its main benefits and recommendations to practice it.
Establishing formal and informal practices	To teach the importance of the establishment not only of formal but also of informal practice.
Through management, body scan practice and values	To help people to see the importance of values to keep a regular mindfulness practice.
Self-compassion. Integrating mindfulness in everyday life	To establish a regular practice of mindfulness to be indefinitely kept.

Instruments

Demographic Variables

We gathered sociodemographic data such as gender, age, place of residence, family status, living with family or alone, level of studies, work status, and income level according to national minimum wage (NMW) as well as clinical variables such as taking psychopharmacological medication (yes vs no) and the number of general practitioners visits in the previous 12 months.

Outcomes

The Spanish version of PHQ-9 [36], as a continuous variable, was used as the primary outcome measure in all 4 assessments, from pretreatment (time 1) to 12-month follow-up (time 4). PHQ-9 is one of the most widely used instruments to evaluate the presence and severity of depressive symptoms. Participants describe their mood according to the last 2 weeks before evaluation. Items range from 0 to 3 denoting *not at all*, *several days*, *more than half the days*, and *nearly every day*, respectively. Total scores range from 0 to 27. The Spanish version has been shown to have good psychometric properties (for the diagnosis of any disorder, k=0.74; overall accuracy, 88%; sensitivity, 87%; and specificity, 88%).

Secondary outcomes included the visual analog scale (VAS) of the EuroQol (*EuroQol—a new facility for the measurement of health-related quality of life*, 1990), in its Spanish version [37]; the Short-Form Health Survey (SF-12) [38], in its Spanish version [39], as a measure of health-related quality of life and functioning; the Positive and Negative Affect Schedule (PANAS) [40], in its Spanish version [41], as a measure of positive and negative affect; and the Pemberton Happiness Index (PHI) [42] as a measure of general well-being.

The VAS is a vertical line on which the best and worst possible health states are scored, 100 or 0, respectively. The SF-12 scoring algorithm yields a physical and mental component scale, and both were used as continuous variables applying Spanish norms. The PANAS evaluates 2 independent dimensions: positive affect and negative affect and were used as continuous variables. *Trait* version was used in this study. The overall PHI index was used as a continuous measure of general well-being, and the scores ranged from 0 to 10.

Sample Size

Required sample size was 240 participants, 60 participants in each condition [32]. This estimation was calculated according to the literature, with a SD of 9.2 and a mean of 16.2 in the iTAU group [43], 14.59 in PAPP group [44], 16.12 in HLP group [45], and 10.3 in the MP group [43], accepting an α of .05 and a β risk<0.2 in a bilateral contrast and assuming a 25.0% (60/240) patient loss to follow-up. This sample size also allows for the calculation of the clinically significant difference in the main outcome variable, PHQ-9 [36], and this difference has been placed at 5 points.

Data Analysis

First, demographic and outcome variables were characterized through descriptive exploratory analysis. Database scrutiny revealed increasing percentages of whole wave missingness in primary and secondary outcome variables along the follow-up. Missingness effects were, thereafter, assessed through sensitivity analyses for demographic variables, intervention groups, and baseline outcomes, considering dropout as study abandonment, with or without subject return, at any assessment period. Although an association between collected variables and study attrition had been detected, no association was reported between outcome values and follow-up missingness; hence, missing at random was assumed for primary and secondary outcome variables. Finally, we implemented Multiple Imputation with Chained Equations (MICE) to replace the outcome missing values, performing 100 imputation models with 100 iterations per model.

We conducted paired *t*-tests and Wilcoxon Signed Rank tests to estimate PHQ-9 primary outcome differences between study stages. Analysis of variance and Tukey's range test were also displayed to examine outcome differences between intervention groups at each time point. In addition, unadjusted and adjusted to sex and age linear regression models were performed. Consecutively, Hedge's (g) effect size index was calculated for each unadjusted regression model. The same approach was used compare secondary results from SF-12 Mental and Physical subscales scores, EuroQol (VAS) scores, and PHI global scores. We used a Complier Average Causal Effect (CACE) analysis to determine the number of completed modules effect on PHQ-9 posttreatment scores, defining compliance as 4 completed modules (100%) and analyzing these same effects per module. In all the analyses, we used a 2-sided test at 5% significance level. Data analyses were implemented with R (3.15.) and Stata 15 (StataCorp).

Results

Baseline Characteristics and Attrition Rates

A total of 221 recruited participants met inclusion criteria and agreed to participate after baseline assessments (Figure 1). The number of recruited participants varied across regions: 75 from Zaragoza, 76 from Málaga, and 70 from Mallorca. A total of 57 participants were designated to iTAU, 54 to HLP, 54 to MP, and the remaining 56 to PAPP. Although all participants did not provide their complete baseline sociodemographic data, no statistically significant differences were found between intervention groups after randomization (Table 2).

Attrition rates increased significantly as study went forward: primary outcome PHQ-9 data were collected for the 72.4% (160/221) of participants at time 1, 57.5% (127/221) at time 2, 46.2% (102/221) at time 3, and 43.9% (97/221) at time 4. Missingness analysis does not report outcome baseline significant differences between dropout and nondropout groups. Conversely, differences in missingness were found between intervention groups, with 28% (15/54) in HLP and 24% (13/54) in MP presenting significative less dropout subjects (*P*<.01) than 45% (25/56) in PAPP and 63% (36/57) in iTAU groups. Baseline age differences were also detected between dropout and nondropout groups at time 2 (*P*=.04) and time 3 (*P*=.03). After MICE, mean depression severity assessed by PHQ-9 score at pretreatment stage (time 1) was 15.33 (SD 5.76) with a median of 15.5, which agrees for a moderate depression level.

Figure 1. Flow diagram.

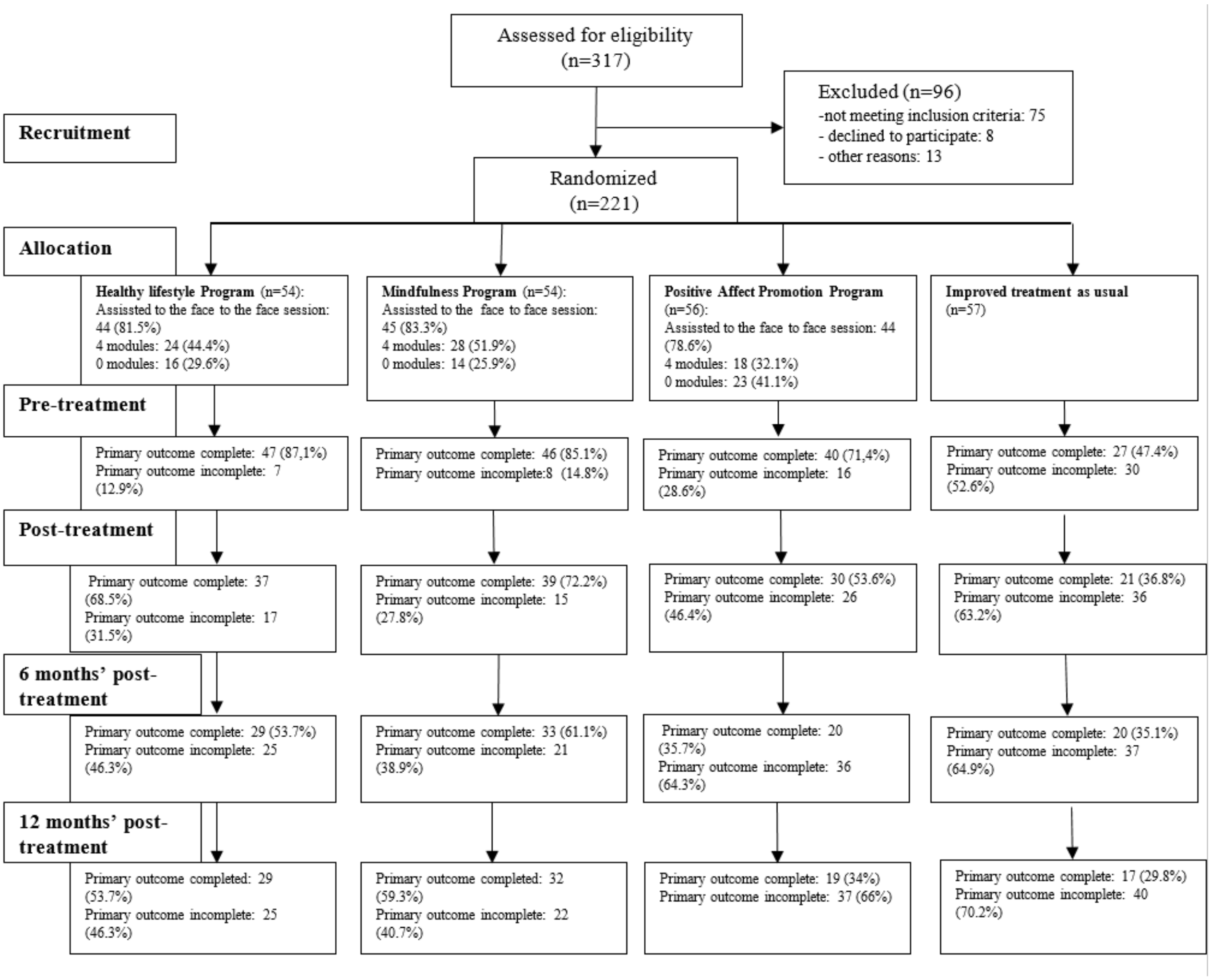


Table 2. Baseline characteristics of participants between intervention groups.

Intervention characteristics and measures	Intervention groups			
	iTAU[a]	HLP[b]	MP[c]	PAPP[d]
Sociodemographic characteristics				
Age, mean (SD)	44.54 (16.10)	44.67 (9.98)	47.50 (13.09)	44.53 (10.23)
Sex (female), n (%)	41 (72)	40 (74)	47 (87)	44 (79)
Married, n (%)	24 (52)	23 (47)	27 (59)	31 (62)
Living with family or couple, n (%)	38 (83)	36 (75)	34 (76)	44 (88)
High education, n (%)	14 (33)	18 (45)	17 (43)	19 (40)
Employed, income level, n (%)	20 (44)	22 (47)	17 (40)	25 (50)
<1 NMW[e]	6 (19)	10 (30)	7 (23)	13 (33)
1-2 NMW	16 (52)	16 (48)	7 (23)	14 (36)
2-4 NMW	8 (26)	7 (21)	12 (40)	12 (31)
>4 NMW	1 (3)	0 (0)	4 (13)	0 (0)
Clinical measures				
Depression severity				
PHQ-9[f], mean (SD); median (IQR)	12.46 (2.10); 13.0 (11-14)	12.57 (2.46); 13.5 (11-14)	12.67 (2.56); 13.0 (11-14)	12.63 (2.03); 13.0 (11.5-14)
Perceived health				
Physical SF-12[g], mean (SD); median (IQR)	43.06 (11.03); 41.38 (35.49-51.89)	42.16 (10.66); 39.84 (35.07-52.02)	42.52 (9.75); 42.11 (35.58-48.27)	45.17 (13.54); 45.74 (35.04-55.65)
Mental SF-12, mean (SD); median (IQR)	26.75 (9.62); 24.54 (20.97-30.58)	27.59 (9.61); 25.98 (21.47-32.75)	26.96 (10.86); 24.51 (20.93-28.66)	26.22 (9.97); 23.79 (19.77-29.85)
VAS EuroQol[h], mean (SD); median (IQR)	51.60 (17.95); 50 (40-60)	52.22 (21.63); 50 (30-70)	48.91 (26.01); 50 (30-70)	49.00 (19.05); 50 (40-60)
Overall PHI[i] index, mean (SD); median (IQR)	4.3 (1.86); 4.33 (2.92-5.42)	4.4 (1.95); 4.42 (3-5.42)	4.32 (1.83); 4.42 (2.96-5.17)	4.25 (1.98); 4.58 (2.83-5.42)
Positive affect PANAS[j], mean (SD); median (IQR)	18.56 (6.43); 17 (14-23)	19.04 (6.94); 18 (13-25)	19.48 (6.60); 18 (15-22)	18.09 (6.00); 17 (13-22)
Negative affect PANAS, mean (SD); median (IQR)	28.27 (8.45); 27 (22-34)	28.91 (8.34); 29 (23-35)	27.89 (8.07); 27 (22-32.75)	29.46 (8.79); 27 (23-35.5)

[a] iTAU: improved treatment as usual.

[b] HLP: healthy lifestyle program.

[c] MP: mindfulness program.

[d] PAPP: positive affect promotion program.

[e] NMW: national minimum wage.

[f] PHQ-9: Patient Health Questionnaire-9 items.

[g] SF-12: 12-item Short-Form Health Survey.

[h] VAS EuroQol: visual analog scale of the EuroQol.

[i] PHI: Pemberton Happiness Index.

[j] PANAS: Positive and Negative Affect Schedule.

Primary Analysis

All our primary and secondary results were extracted from databases with high rates of attrition, 44.7% (395/884), which were thereafter imputed. Thus, the following results should be considered more as suggestive hypothesis rather than empirical statements. To this extent, we found significant decreases of PHQ-9 scores (*P*<.001) in all interventions (iTAU included) from pretreatment (mean 15.33, SD 5.76) to posttreatment (mean 10.19, SD 6.42) and up to 6 and 12 months (time 3: mean 9.39, SD 6.59 and time 4: mean 9.68, SD 6.14). After treatment, a moderate decrease was detected in PHQ scores from HLP and MP relative to iTAU: iTAU versus HLP (β=–3.05; *P*=.01) and iTAU versus MP (β=–3.00; *P*=.01; Table 3). In contrast, we observed no significant PHQ-9 differences between PAPP and iTAU, nor between any psychotherapeutic strategies, throughout the study (Table 3). Adjusted regression models replicated these same findings with poor variation (Multimedia Appendix 1). CACE analysis reported posttreatment dose-response significant decrease in PHQ-9 scores in HLP and PAPP in both imputed and adjusted models (Table 4, Multimedia Appendix 2). Despite the fact that the compliance effects (4 modules vs >4 modules) tended to disappear in the long term, the effects per module remained conserved (Table 4).

Table 3. Primary outcome analysis with imputed data (N=221): intervention comparisons along the follow-up.

Primary outcome	Time 1 (pretreatment)	Time 2 (posttreatment)	Time 3 (6 months)	Time 4 (12 months)
PHQ-9[a]				
iTAU[b] vs HLP[c]				
g[d]	0.25	–0.50	–0.22	–0.06
P value	.20	*.01* [e]	.23	.73
β[f] (95% CI)	1.41 (–0.75 to 3.57)	–3.05 (–5.43 to –0.68)	–1.52 (–3.98 to 0.95)	–0.40 (–2.71 to 1.91)
iTAU vs MP[g]				
g	–0.17	0.47	0.24	–0.01
P value	.36	*.01*	.18	.96
β (95% CI)	1.00 (–1.16 to 3.16)	–3.00 (–5.37 to –0.63)	–1.68 (–4.15 to 0.78)	0.06 (–2.25 to 2.37)
iTAU vs PAPP[h]				
g	–0.23	0.23	0.31	0.01
P value	.20	.22	.10	.99
β (95% CI)	1.40 (–0.74 to 3.54)	–1.46 (–3.81 to 0.89)	–2.08 (–4.52 to 0.37)	–0.02 (–2.31 to 2.27)
HLP vs MP				
g	0.07	-0.01	0.03	–0.07
P value	.71	.96	.90	.70
β (95% CI)	0.41 (–1.78 to 2.57)	–0.06 (–2.46 to 2.35)	0.17 (–2.33 to 2.67)	–0.46 (–2.81 to 1.88)
HLP vs PAPP				
g	0.00	–0.25	0.09	–0.06
P value	.99	.19	.66	.75
β (95% CI)	0.01 (–2.16 to 2.18)	–1.59 (–3.97 to 0.79)	0.56 (–1.92 to 3.04)	–0.38 (–2.70 to 1.94)
MP vs PAPP				
g	–0.07	–0.23	0.06	0.01
P value	.72	.21	.76	.94
β (95% CI)	0.40 (–1.77 to 2.57)	1.53 (–0.85 to 3.92)	–0.39 (–2.87 to 2.08)	–0.08 (–2.41 to 2.24)

[a] PHQ-9: Patient Health Questionnaire-9 items.

[b] iTAU: improved treatment as usual.

[c] HLP: healthy lifestyle program.

[d] g: Hedge’s effect size measure.

[e] Statistically significant values (*P*<.05) are shown in italics.

[f] β: regression coefficient.

[g] MP: mindfulness program.

[h] PAPP: positive affect promotion program.

Table 4. Dose-response in imputed primary outcome at posttreatment and along the follow-up.

Interventions	Pretreatment to posttreatment		Pretreatment to posttreatment		Pretreatment to posttreatment	
	β[a] (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
HLP[b]						
CACE[c] analysis[d]	–3.32 (4.43 to –2.21)	*.004* [e]	–0.51 (2.03 to 1.02)	.74	–1.03 (2.61 to 0.56)	.52
Effect per session	–0.28 (0.34 to –0.21)	*.001*	–0.17 (0.25 to –0.09)	*.03*	–0.06 (0.13 to 0.02)	.45
MP[f]						
CACE analysis	0.93 (0.71 to 2.57)	.57	1.01 (0.72 to 2.74)	.56	1.13 (0.5 to 2.76)	.49
Effect per session	0.26 (0.22 to 0.74)	.59	–0.17 (0.25 to –0.09)	*.03*	–0.06 (0.13 to 0.02)	.45
PAPP[g]						
CACE analysis	0.24 (1.30 to 1.78)	.88	0.96 (0.66 to 2.57)	.56	0.30 (1.24 to 1.84)	.85
Effect per session	–0.28 (0.34 to –0.21)	*.001*	–0.17 (0.25 to –0.09)	*.03*	–0.06 (0.13 to 0.02)	.45
All						
CACE analysis	–2.11 (2.86 to –1.36)	*.006*	–0.49 (1.37 to 0.4)	.58	–0.22 (1.04 to 0.6)	.79
Effect per session	–0.28 (0.34 to –0.21)	*.001*	–0.17 (0.25 to –0.09)	*.03*	–0.06 (0.13 to 0.02)	.45

[a] β: regression coefficients.

[b] HLP: healthy lifestyle program.

[c] CACE: Complier Average Causal Effect.

[d] Compliance as attendance >4 modules.

[e] Statistically significant values (*P*<.05) are shown in italics.

[f] MP: mindfulness program.

[g] PAPP: positive affect promotion program..

Secondary Analysis

Imputed Mental and Physical SF-12 scores significantly increased in all intervention groups (iTAU included) from pretreatment to posttreatment (Mental SF-12: *P*<.001; Physical SF-12: *P*<.001) and from posttreatment to 6 months after treatment (Mental SF-12: *P*<.001; Physical SF-12: *P*=.02). Although differences between these intervention groups disappeared in the long term (Table 5), iTAU Mental SF-12 scores were higher than HLP scores (β=–5.32; *P*=.02) and PAPP scores (β=–7.72; *P*=.001) at posttreatment. Conversely, we determined posttreatment increases in HLP Physical SF-12 scores relative to iTAU (β=4.58; *P*=.047) and from MP group compared with iTAU (β=5.32; *P*=.02). Physical SF-12 HLP, MP, and PAPP group differences relative to iTAU were observed up to 6 months after treatment (Table 5). These results were replicated similarly by adjusted regression coefficients (Multimedia Appendix 3).

Although EuroQol (VAS) significant differences were detected from pretreatment to 6 months (*P*<.001) and up to 12 months after treatment (*P*<.001) in all intervention groups, no meaningful differences among the groups were observed at any time (Table 6). Otherwise, PHI scores rose significantly at 6 and 12-months after treatment relative to pretreatment (*P*<.001), and all psychotherapy interventions, except HLP, reported better PHI results than iTAU treatment at time 3 (Table 6). These results were replicated similarly by adjusted regression coefficients (Multimedia Appendix 4). PANAS negative affect scale decreased in all intervention groups at posttreatment (*P*<.001), at 6 months (*P*<.001), and at 12 months (*P*<.001) compared with pretreatment. In contrast, all PANAS positive affect scores increased significantly throughout study time relative to pretreatment (*P*<.001). However, we found no significant PANAS positive affect differences between intervention groups throughout the study (Table 7). Regarding PANAS negative scale, only PAPP intervention was significantly lower than iTAU when were compared 12 months after treatment (Table 7). These results were replicated similarly by adjusted regression coefficients (Multimedia Appendix 5).

Table 5. Short-Form Health Survey-12 (Mental and Physical) outcome analysis with imputed data (N=221): intervention comparisons along the follow-up.

Secondary outcomes	Time 1 (pretreatment)	Time 2 (posttreatment)	Time 3 (6 months)	Time 4 (12 months)
Mental Scale SF-12[a]				
iTAU[b] vs HLP[c]				
g[d]	–0.06	–0.42	–0.11	0.01
P value	.73	*.02*[e]	.53	.98
β[f] (95% CI)	–0.63 (–4.25 to 2.99)	–5.32 (–9.91 to –0.72)	–1.53 (–6.30 to 3.23)	–0.05 (–5.04 to 4.93)
iTAU vs MP[g]				
g	0.14	0.16	0.01	–0.05
P value	.46	.38	.94	.81
β (95% CI)	–1.37 (–4.99 to 2.24)	–2.06 (–6.66 to 2.53)	–0.18 (–4.95 to 4.58)	0.62 (–4.36 to 5.61)
iTAU vs PAPP[h]				
g	0.15	0.67	0.22	0.24
P value	.45	*.001*	.28	.20
β (95% CI)	–1.37 (–4.96 to 2.21)	–7.72 (–12.27 to –3.16)	–2.61 (–7.33 to 2.11)	–3.2 (–8.14 to 1.74)
HLP vs MP				
g	0.07	–0.25	–0.1	–0.05
P value	.69	.17	.58	.79
β (95% CI)	–0.74 (–4.41 to 2.93)	3.25 (–1.41 to 7.91)	1.35 (–3.48 to 6.18)	0.68 (–4.37 to 5.73)
HLP vs PAPP				
g	0.01	0.17	–0.07	–0.12
P value	.97	.38	.72	.49
β (95% CI)	–0.07 (–4.22 to 4.08)	–2.04 (–6.59 to 2.51)	0.76 (–3.41 to 4.92)	1.36 (–2.51 to 5.24)
MP vs PAPP				
g	0.23	0.23	0.25	0.09
P value	.20	.23	.23	.62
β (95% CI)	–2.72 (–6.87 to 1.43)	–2.77 (–7.32 to 1.78)	–2.53 (–6.69 to 1.64)	–0.97 (–4.85 to 2.9)
Physical Scale SF-12				
iTAU vs HLP				
g	0.04	0.37	0.39	–0.19
P value	.86	*.047*	*.03*	.35
β (95% CI)	0.36 (–3.77 to 4.5)	4.58 (0.05 to 9.11)	4.73 (0.58 to 8.87)	–1.85 (–5.7 to 2.01)
iTAU vs MP				
g	–0.25	–0.42	–0.73	–0.05
P value	.15	*.02*	*.001*	.80
β (95% CI)	3.01 (–1.12 to 7.14)	5.32 (0.79 to 9.85)	8.01 (3.87 to 12.16)	0.49 (–3.37 to 4.35)
iTAU vs PAPP				
g	–0.03	–0.21	–0.50	0.05
P value	.89	.27	*.009*	.80
β (95% CI)	0.29 (–3.8 to 4.39)	2.54 (–1.94 to 7.03)	5.49 (1.38 to 9.6)	–0.48 (–4.3 to 3.34)
HLP vs MP				
g	–0.22	–0.06	–0.29	–0.21

Secondary outcomes	Time 1 (pretreatment)	Time 2 (posttreatment)	Time 3 (6 months)	Time 4 (12 months)
P value	.21	.75	.13	.24
β (95% CI)	2.65 (–1.54 to 6.83)	0.73 (–3.86 to 5.32)	3.28 (–0.92 to 7.49)	2.34 (–1.57 to 6.25)
HLP vs PAPP				
g	0.08	0.20	0.09	0.23
P value	.69	.31	.66	.22
β (95% CI)	–0.74 (–4.37 to 2.9)	–2.4 (–7.02 to 2.22)	–1.08 (–5.86 to 3.71)	–3.15 (–8.15 to 1.86)
MP vs PAPP				
g	0.01	0.47	0.20	0.28
P value	>.99	*.02*	.32	.13
β (95% CI)	0.01 (–3.63 to 3.64)	–5.65 (–10.27 to –1.03)	–2.43 (–7.21 to 2.36)	–3.82 (–8.83 to 1.18)

[a] SF-12: 12-item Short-Form Health Survey.

[b] iTAU: improved treatment as usual.

[c] HLP: healthy lifestyle program.

[d] g: Hedge’s effect size measure.

[e] Statistically significant values (*P*<.05) are shown in italics.

[f] β: regression coefficient.

[g] MP: mindfulness program.

[h] PAPP: positive affect promotion program.

Table 6. Visual analog scale of the EuroQol and Pemberton Happiness Index outcome analysis with imputed data (N=221): intervention comparisons along the follow-up.

Secondary outcomes	Time 1 (pretreatment)	Time 3 (6 months)	Time 4 (12 months)
VAS EuroQol[a]			
iTAU[b] vs HLP[c]			
g[d]	0.02	–0.20	–0.33
P value	.91	.31	.08
β[e] (95% CI)	0.47 (–7.93 to 8.87)	–3.82 (–11.17 to 3.53)	–6.56 (–13.86 to 0.75)
iTAU vs MP[f]			
g	0.17	0.18	0.16
P value	.33	.33	.39
β (95% CI)	–4.16 (–12.56 to 4.24)	–3.63 (–10.98 to 3.72)	–3.17 (–10.47 to 4.14)
iTAU vs PAPP[g]			
g	0.18	0.10	0.09
P value	.36	.60	.63
β (95% CI)	–3.9 (–12.22 to 4.43)	–1.95 (–9.23 to 5.33)	–1.76 (–9.00 to 5.48)
HLP vs MP			
g	0.19	–0.01	–0.17
P value	.29	.96	.58
β (95% CI)	–4.63 (–13.14 to 3.88)	0.19 (–7.26 to 7.63)	–2.04 (–9.21 to 5.2)
HLP vs PAPP			
g	0.21	–0.10	–0.24
P value	.31	.62	.20
β (95% CI)	–4.37 (–12.8 to 4.07)	1.86 (–5.52 to 9.24)	4.79 (–2.55 to 12.13)
MP vs PAPP			
g	–0.01	–0.08	–0.07
P value	.95	.65	.71
β (95% CI)	0.26 (–8.17 to 8.7)	1.68 (–5.7 to 9.06)	1.4 (–5.93 to 8.74)
PHI[h]			
iTAU vs HLP			
g	–0.05	0.19	–0.18
P value	.81	.30	.33
β (95% CI)	–0.09 (–0.79 to 0.61)	0.40 (–0.35 to 1.16)	–0.32 (–0.98 to 0.33)
iTAU vs MP			
g	0.08	–0.49	–0.06
P value	.65	*.01*[i]	.75
β (95% CI)	–0.16 (–0.86 to 0.54)	0.98 (0.22 to 1.73)	0.11 (–0.55 to 0.77)
iTAU vs PAPP			
g	0.10	–0.42	0.09
P value	.59	*.03*	.61
β (95% CI)	–0.19 (–0.89 to 0.5)	0.82 (0.07 to 1.57)	–0.17 (–0.82 to 0.48)
HLP vs MP			
g	0.04	–0.27	–0.25

Secondary outcomes	Time 1 (pretreatment)	Time 3 (6 months)	Time 4 (12 months)
P value	.84	.14	.20
β (95% CI)	–0.07 (–0.78 to 0.64)	0.57 (–0.19 to 1.34)	0.43 (–0.23 to 1.1)
HLP vs PAPP			
g	0.06	–0.2	–0.08
P value	.77	.28	.65
β (95% CI)	–0.11 (–0.81 to 0.6)	0.42 (–0.34 to 1.17)	0.15 (–0.51 to 0.81)
MP vs PAPP			
g	0.02	0.08	0.16
P value	.93	.68	.41
β (95% CI)	–0.03 (–0.74 to 0.67)	–0.16 (–0.92 to 0.6)	–0.28 (–0.94 to 0.38)

[a] VAS EuroQol: visual analog scale of the EuroQol.

[b] iTAU: improved treatment as usual.

[c] HLP: healthy lifestyle program.

[d] g: Hedge’s effect size measure.

[e] β: regression coefficient.

[f] MP: mindfulness program.

[g] PAPP: positive affect promotion program.

[h] PHI: Pemberton Happiness Index.

[i] Statistically significant values (*P*<.05) are shown in italics.

Table 7. Positive and Negative Affect Scales outcome analysis with imputed data (N=221): intervention comparisons along the follow-up.

Secondary outcomes	Time 1 (pretreatment)	Time 2 (posttreatment)	Time 3 (6 months)	Time 4 (12 months)
Positive Scale PANAS[a]				
iTAU[b] vs HLP[c]				
g[d]	0.07	–0.01	0.25	0.34
P value	.72	.94	.17	.07
β[e] (95% CI)	0.45 (–1.96 to 2.85)	–0.12 (–3.47 to 3.23)	2.24 (–1.00 to 5.48)	3.36 (–0.30 to 7.01)
iTAU vs MP[f]				
g	0.14	0.00	–0.16	–0.33
P value	.44	>.99	.41	.09
β (95% CI)	–0.94 (–3.35 to 1.46)	–0.01 (–3.36 to 3.34)	1.35 (–1.88 to 4.59)	3.19 (–0.46 to 6.85)
iTAU vs PAPP[g]				
g	0.21	0.31	0.05	–0.15
P value	.25	.10	.81	.41
β (95% CI)	–1.41 (–3.79 to 0.97)	–2.78 (–6.09 to 0.54)	–0.40 (–3.61 to 2.81)	1.51 (–2.11 to 5.13)
HLP vs MP				
g	0.22	–0.01	0.10	0.02
P value	.26	.95	.59	.93
β (95% CI)	–1.39 (–3.83 to 1.05)	0.11 (–3.28 to 3.5)	–0.89 (–4.17 to 2.39)	–0.17 (–3.87 to 3.54)
HLP vs PAPP				
g	0.29	0.31	0.3	0.19
P value	.13	.12	.11	.32
β (95% CI)	–1.86 (–4.27 to 0.56)	–2.66 (–6.02 to 0.71)	–2.64 (–5.89 to 0.61)	–1.85 (–5.52 to 1.82)
MP vs PAPP				
g	0.08	0.33	0.21	0.17
P value	.70	.11	.29	.37
β (95% CI)	–0.47 (–2.88 to 1.95)	–2.77 (–6.13 to 0.59)	–1.75 (–5.00 to 1.5)	–1.68 (–5.35 to 1.99)
Negative Scale PANAS[a]				
iTAU vs HLP				
g	–0.12	–0.32	–0.16	–0.36
P value	.52	.09	.39	.07
β (95% CI)	–1.02 (–4.18 to 2.14)	–2.81 (–6.04 to 0.42)	–1.37 (–4.51 to 1.78)	–3.15 (–6.51 to 0.21)
iTAU vs MP				
g	–0.06	0.31	0.16	0.24
P value	.73	.10	.40	.18
β (95% CI)	0.55 (–2.61 to 3.71)	–2.73 (–5.96 to 0.5)	–1.35 (–4.49 to 1.80)	–2.32 (–5.67 to 1.04)
iTAU vs PAPP				
g	0.25	0.31	0.33	0.41
P value	.19	.11	.11	*.03* [h]
β (95% CI)	–2.09 (–5.22 to 1.04)	–2.59 (–5.79 to 0.61)	–2.54 (–5.65 to 0.57)	–3.63 (–6.96 to –0.31)
HLP vs MP				
g	–0.19	–0.01	0.00	–0.09
P value	.33	.96	.99	.63

Secondary outcomes	Time 1 (pretreatment)	Time 2 (posttreatment)	Time 3 (6 months)	Time 4 (12 months)
β (95% CI)	1.57 (–1.63 to 4.78)	0.07 (–3.20 to 3.35)	0.02 (–3.17 to 3.20)	0.83 (–2.57 to 4.23)
HLP vs PAPP				
g	0.13	–0.03	0.14	0.06
P value	.51	.90	.46	.78
β (95% CI)	–1.07 (–4.24 to 2.11)	0.21 (–3.03 to 3.46)	–1.17 (–4.33 to 1.98)	–0.48 (–3.85 to 2.89)
MP vs PAPP				
g	0.3	–0.02	0.15	0.14
P value	.10	.93	.46	.44
β (95% CI)	–2.64 (–5.82 to 0.53)	0.14 (–3.10 to 3.38)	–1.19 (–4.35 to 1.96)	–1.32 (–4.69 to 2.06)

[a] PANAS: Positive and Negative Affect Schedule.

[b] iTAU: improved treatment as usual.

[c] HLP: healthy lifestyle program.

[d] g: Hedge’s effect size measure.

[e] β: regression coefficient.

[f] MP: mindfulness program.

[g] PAPP: positive affect promotion program.

[h] Statistically significant values (*P*<.05) are shown in italics.

Internet-Based Program Usage

A total of 81.1% (133/164) participants attended the initial face-to-face session. In HLP, 44% (24/54) of participants completed all web-based modules, in MP, 52% (28/54), and in PAPP, 32% (18/56; χ^2_2=4.4; *P*=.11). In HLP, the median number of sessions completed was 2 (range: 0-4). In MP, the median number of sessions completed was 4 (range: 0-4), and in PAPP, 2 (range: 0-4). There were no significant differences between intervention groups in terms of sessions completed (F_2=1.775; *P*=.17).

Discussion

Principal Findings

The main objective of our study was to examine the efficacy of 3 low-intensity, internet-based psychological interventions when compared with that of the control condition (iTAU) in primary care in Spain.

Our main finding was that there were differences in the short term in favor of internet-based psychological interventions, specifically HLP and MP. This finding is consistent with the literature that has shown that brief psychotherapy is efficacious for the treatment of depression in primary care [12] and is also in line with the previous literature that has demonstrated that internet-based intervention programs are effective for the treatment of depression [18,46-48]. However, no differences were found in depression severity between PAPP and iTAU. This finding is not in line with the previous literature that has shown that this psychotherapy is effective in reducing depression [49]. Moreover, there were no differences found in the medium and long term between intervention and control groups. This result differs from findings of previous systematic reviews and meta-analysis that have evidenced that electronic health interventions effectively reduce depressive symptoms [11,50,51]. However, differences between populations may affect the generalizability of these studies [52]. It has been argued that psychotherapy may be less effective in primary care than in other settings, mainly due to the considerable heterogeneity in the severity of symptoms and the lower motivation among patients in primary care settings [53]. Our findings also show the faster effect of our interventions, with moderate effect size at posttreatment. The lack of mid- and long-term efficacy could be explained by a floor effect. Care as usual is not a clearly defined treatment, entailing different interventions [52]. Given that the usual care was improved before the interventions and that pharmacological treatment were offered to all groups, no major effects can be expected when web-based interventions were given. Our results also show that there were no differences between interventions throughout the study regarding depressive symptomatology. This finding is in line with that of the study by Cuijpers et al [54] who found that all therapies are equally effective in the treatment of depression.

Furthermore, we found differences between the 3 interventions groups and control group (iTAU) regarding health-related quality of life. In particular, we observed short- and medium-term differences in favor of HLP and MP in physical health status, and in medium term for PAPP. These results demonstrated that the 3 low-intensity, internet-based psychological interventions in primary care are also effective in improving medium- and long-term quality of life. This finding is highly important, for, as is well known, depression is associated with serious disability and loss in quality of life [55,56], suggesting that it could be a useful tool to improve quality of life of patients with depression.

Differences were also found in well-being and affect between intervention and control groups. In particular, we observed differences in medium term regarding well-being in MP compared with that in iTAU. In PAPP, differences were found in medium and long term with regard to well-being and, in long term, with regard to negative affect in favor of psychotherapy. This finding shows us that although PAPP seems to be ineffective in reducing depressive symptomatology, it could be effective in improving health-related quality of life and well-being in patients with depression.

Regarding treatment adherence, our completion treatment rates were relatively low in each intervention group. Dropout treatment is common in internet intervention programs, although rates vary depending on the support provided along the intervention or the context [47]. Regarding primary care, our adherence rates are higher than those reported in a previous study, in which participants completing all treatment sessions did not exceed 20% [57].

A possible explanation could be that in our study, there was an initial face-to-face group session, in which the final goal was to reinforce commitment and adherence to treatment, as well as, to explain the program structure and main components of treatment, clarify the instructions for the use of the web-based platform, and motivate participants to change. Thus, perhaps these measures could have increased our completion rates.

Limitations and Strengths

This trial presents several limitations, which should be mentioned. First, not all participants completed posttest measurements, and a high attrition rate at follow-up was found. Although missing values were corrected by using multiple imputations, the results should be interpreted with caution. Second, just as difficulties in recruiting patients is an important issue in clinical trials [58-60], general practitioners may also experience problems in recruiting patients owing to their overload schedule, and our sample size is slightly lower than the expected. Finally, treatment directed to depression problems with the general practitioners in the iTAU group was not recorded. It would be necessary to consider these variables in the future to analyze possible influences on between-group results.

Our study has a significant strength: to the best of our knowledge, this is the first trial in Spain aimed at improving the symptomatology and quality of life of patients with depression using low-intensity interventions applied by the information and communication technologies. The treatment programs used in this study include therapeutic strategies based on mindfulness, healthy lifestyle, and positive affect, which have proven their efficacy for the treatment of depression; nevertheless, it is still the first study that adapts these interventions to information and communication technologies.

Conclusions

This study has 3 important conclusions. First, 2 low-intensity, internet-based psychological interventions (HLP and MP) for the treatment of depression in primary care were more effective than iTAU at posttreatment. Second, all low-intensity, internet-based psychological interventions were also effective in improving medium- and long-term quality of life. Finally, PAPP was effective for improving health-related quality of life and well-being in patients with depression. Nevertheless, it is important to examine possible reasons that could be implicated in the ineffectiveness of PAPP in reducing depressive symptomatology, such as the intervention length, population, and treatment adherence, to increase its effectiveness in future studies of internet-based interventions programs for depression. Overall, our results suggest that although low-intensity, internet-based psychological programs are an efficacious therapeutic option for the treatment of depression in primary care, subsequent and more complex analyses are necessary to explain the reasons why some interventions appeared to affect some outcomes but not others. Furthermore, more research is still needed to assess the cost-effectiveness analysis of these interventions.

Acknowledgments

This study was financed by the Instituto de Salud Carlos III of the Spanish Ministry of Economy and Competitiveness with the PI13/01171 grant (Efficacy and cost-effectiveness of the low-intensity psychological interventions applied by ICTs for the treatment of depression in primary care: a controlled trial). The authors would like to thank the Institut d'Investigació Sanitaria de les Illes Balears (IdiSBa) for the assistance received on statistical analysis queries.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Primary outcome analysis with imputed data adjusted to Sex and Age (N=221): intervention comparisons along the follow-up.

[DOCX File , 16 KB - jmir_v22i6e15845_app1.docx]

Multimedia Appendix 2

Dose-response in imputed and adjusted (Sex and Age) primary outcome at post-treatment and along the follow-up.

[DOCX File , 16 KB - jmir_v22i6e15845_app2.docx]

Multimedia Appendix 3

SF-12 (Mental and Physical) analysis with imputed data adjusted to Sex and Age (N=221): intervention comparisons along the follow-up.
[DOCX File , 19 KB - jmir_v22i6e15845_app3.docx]

Multimedia Appendix 4
EuroQoL (VAS) and PHI analysis with imputed data adjusted to Sex and Age (N=221): intervention comparisons along the follow-up.
[DOCX File , 18 KB - jmir_v22i6e15845_app4.docx]

Multimedia Appendix 5
PANAS (positive and negative affect scales) analysis with imputed data adjusted to Sex and Age (N=221): intervention comparisons along the follow-up.
[DOCX File , 19 KB - jmir_v22i6e15845_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1591 KB - jmir_v22i6e15845_app6.pdf]

References

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
2. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012 Mar;35(1):1-14 [FREE Full text] [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
3. Fernández CF, Martínez M, Martín AB, Pérez-Fuentes MD, Jurado MD, Linares JJ. Prevalencia de la depresión en España: Análisis de los últimos 15 años. Eur J Investig Health Psychol Educ 2015;5(2):267. [doi: 10.30552/ejihpe.v5i2.118]
4. Pence BW, O'Donnell JK, Gaynes BN. The depression treatment cascade in primary care: a public health perspective. Curr Psychiatry Rep 2012 Aug;14(4):328-335 [FREE Full text] [doi: 10.1007/s11920-012-0274-y] [Medline: 22580833]
5. Craven MA, Bland R. Depression in primary care: current and future challenges. Can J Psychiatry 2013 Aug;58(8):442-448. [doi: 10.1177/070674371305800802] [Medline: 23972105]
6. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). J Am Med Assoc 2003 Jun 18;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
7. Verhaak PF, Prins MA, Spreeuwenberg P, Draisma S, van Balkom TJ, Bensing JM, et al. Receiving treatment for common mental disorders. Gen Hosp Psychiatry 2009;31(1):46-55. [doi: 10.1016/j.genhosppsych.2008.09.011] [Medline: 19134510]
8. Barley EA, Murray J, Walters P, Tylee A. Managing depression in primary care: A meta-synthesis of qualitative and quantitative research from the UK to identify barriers and facilitators. BMC Fam Pract 2011 Jun 9;12:47 [FREE Full text] [doi: 10.1186/1471-2296-12-47] [Medline: 21658214]
9. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006 Dec;32(3):254-258. [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
10. Heddaeus D, Steinmann M, Daubmann A, Härter M, Watzke B. Treatment selection and treatment initialization in guideline-based stepped and collaborative care for depression. PLoS One 2018;13(12):e0208882 [FREE Full text] [doi: 10.1371/journal.pone.0208882] [Medline: 30586371]
11. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
12. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: A systematic review. J Affect Disord 2018 Jan 1;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]
13. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
14. Nieuwsma JA, Trivedi RB, McDuffie J, Kronish I, Benjamin D, Williams JW. Brief psychotherapy for depression: a systematic review and meta-analysis. Int J Psychiatry Med 2012;43(2):129-151 [FREE Full text] [doi: 10.2190/PM.43.2.c] [Medline: 22849036]

15. Cape J, Whittington C, Buszewicz M, Wallace P, Underwood L. Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. BMC Med 2010 Jun 25;8:38 [FREE Full text] [doi: 10.1186/1741-7015-8-38] [Medline: 20579335]
16. Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. Psychol Sci Public Interest 2002 Nov;3(2):39-77. [doi: 10.1111/1529-1006.00008] [Medline: 26151569]
17. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
18. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in Spain: a randomized controlled trial. J Med Internet Res 2016 Aug 26;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
19. Görges F, Oehler C, von Hirschhausen E, Hegerl U, Rummel-Kluge C. GET.HAPPY - Acceptance of an internet-based self-management positive psychology intervention for adult primary care patients with mild and moderate depression or dysthymia: A pilot study. Internet Interv 2018 Jun;12:26-35 [FREE Full text] [doi: 10.1016/j.invent.2018.03.001] [Medline: 30135766]
20. Wahbeh H, Oken BS. Internet mindfulness meditation intervention for the general public: pilot randomized controlled trial. JMIR Ment Health 2016 Aug 8;3(3):e37 [FREE Full text] [doi: 10.2196/mental.5900] [Medline: 27502759]
21. Wahbeh H. Internet mindfulness meditation intervention (IMMI) improves depression symptoms in older adults. Medicines (Basel) 2018 Nov 2;5(4):pii: E119 [FREE Full text] [doi: 10.3390/medicines5040119] [Medline: 30400211]
22. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
23. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013 Feb 8;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
24. Chaves C, Lopez-Gomez I, Hervas G, Vazquez C. A comparative study on the efficacy of a positive psychology intervention and a cognitive behavioral therapy for clinical depression. Cogn Ther Res 2017;41(3):417-433 [FREE Full text] [doi: 10.1007/s10608-016-9778-9]
25. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
26. Addington EL, Cheung EO, Bassett SM, Kwok I, Schuette SA, Shiu E, et al. The MARIGOLD study: Feasibility and enhancement of an online intervention to improve emotion regulation in people with elevated depressive symptoms. J Affect Disord 2019 Oct 1;257:352-364. [doi: 10.1016/j.jad.2019.07.049] [Medline: 31302525]
27. Segal ZV, Williams JM, Teasdale JD. Mindfulness-Based Cognitive Therapy for Depression. New York, NY: The Guilford Press; 2012.
28. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
29. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
30. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]
31. Young CL, Trapani K, Dawson S, O'Neil A, Kay-Lambkin F, Berk M, et al. Efficacy of online lifestyle interventions targeting lifestyle behaviour change in depressed populations: A systematic review. Aust N Z J Psychiatry 2018 Sep;52(9):834-846. [doi: 10.1177/0004867418788659] [Medline: 30052063]
32. Castro A, García-Palacios A, García-Campayo J, Mayoral F, Botella C, García-Herrera JM, et al. Efficacy of low-intensity psychological intervention applied by ICTs for the treatment of depression in primary care: a controlled trial. BMC Psychiatry 2015 May 7;15:106 [FREE Full text] [doi: 10.1186/s12888-015-0475-0] [Medline: 25948287]
33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
34. Fernández A, Haro JM, Codony M, Vilagut G, Martínez-Alonso M, Autonell J, et al. Treatment adequacy of anxiety and depressive disorders: primary versus specialised care in Spain. J Affect Disord 2006 Nov;96(1-2):9-20. [doi: 10.1016/j.jad.2006.05.005] [Medline: 16793140]
35. García-Herrera J, Nogueras-Morillas V, Muñoz-Cobos F, Morales-Asensio J. Guía de Práctica Clínica para el tratamiento de la depresión en Atención Primaria. Málaga: GAEDAP; 2011.
36. Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001;63(4):679-686. [doi: 10.1097/00006842-200107000-00021] [Medline: 11485122]

37. Badia X, Roset M, Montserrat S, Herdman M, Segura A. [The Spanish version of EuroQol: a description and its applications. European Quality of Life scale]. Med Clin (Barc) 1999;112(Suppl 1):79-85. [Medline: 10618804]
38. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
39. Vilagut G, Valderas JM, Ferrer M, Garin O, López-García E, Alonso J. [Interpretation of SF-36 and SF-12 questionnaires in Spain: physical and mental components]. Med Clin (Barc) 2008 May 24;130(19):726-735. [doi: 10.1157/13121076] [Medline: 18570798]
40. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
41. Sandín B, Chorot P, Lostao L, Joiner TE, Santed MA, Valiente RM. The PANAS scales of positive and negative affect: Factor analytic validation and cross-cultural convergence. Psicothema 1999;11(1):37-51 [FREE Full text]
42. Hervás G, Vázquez C. Construction and validation of a measure of integrative well-being in seven languages: the Pemberton Happiness Index. Health Qual Life Outcomes 2013 Apr 22;11:66 [FREE Full text] [doi: 10.1186/1477-7525-11-66] [Medline: 23607679]
43. van Aalderen JR, Donders AR, Giommi F, Spinhoven P, Barendregt HP, Speckens AE. The efficacy of mindfulness-based cognitive therapy in recurrent depressed patients with and without a current depressive episode: a randomized controlled trial. Psychol Med 2012 May;42(5):989-1001. [doi: 10.1017/S0033291711002054] [Medline: 22017808]
44. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs technician assistance. PLoS One 2010 Jun 8;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
45. Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Haase AM, Taylor AH, et al. Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. Br Med J 2012 Jun 6;344:e2758 [FREE Full text] [doi: 10.1136/bmj.e2758] [Medline: 22674921]
46. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
47. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
48. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
49. Craske MG, Meuret AE, Ritz T, Treanor M, Dour H, Rosenfield D. Positive affect treatment for depression and anxiety: A randomized clinical trial for a core feature of anhedonia. J Consult Clin Psychol 2019 May;87(5):457-471. [doi: 10.1037/ccp0000396] [Medline: 30998048]
50. Eells TD, Barrett MS, Wright JH, Thase M. Computer-assisted cognitive-behavior therapy for depression. Psychotherapy (Chic) 2014 Jun;51(2):191-197. [doi: 10.1037/a0032406] [Medline: 24059735]
51. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. J Affect Disord 2017 Dec 1;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
52. Massoudi B, Holvast F, Bockting CL, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
53. Cuijpers P, van Straten A, van Schaik A, Andersson G. Psychological treatment of depression in primary care: a meta-analysis. Br J Gen Pract 2009 Feb;59(559):e51-e60 [FREE Full text] [doi: 10.3399/bjgp09X395139] [Medline: 19192368]
54. Cuijpers P. Psychotherapies for adult depression: recent developments. Curr Opin Psychiatry 2015 Jan;28(1):24-29. [doi: 10.1097/YCO.0000000000000121] [Medline: 25415495]
55. Cuijpers P, de Graaf R, van Dorsselaer S. Minor depression: risk profiles, functional disability, health care use and risk of developing major depression. J Affect Disord 2004 Apr;79(1-3):71-79. [doi: 10.1016/S0165-0327(02)00348-8] [Medline: 15023482]
56. van Schaik DJ, Klijn AF, van Hout HP, van Marwijk HW, Beekman AT, de Haan M, et al. Patients' preferences in the treatment of depressive disorder in primary care. Gen Hosp Psychiatry 2004;26(3):184-189. [doi: 10.1016/j.genhosppsych.2003.12.001] [Medline: 15121346]
57. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
58. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun 21;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]

59. de Graaf LE, Gerhards SA, Arntz A, Riper H, Metsemakers JF, Evers SM, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2009 Jul;195(1):73-80. [doi: 10.1192/bjp.bp.108.054429] [Medline: 19567900]
60. Mead N, MacDonald W, Bower P, Lovell K, Richards D, Roberts C, et al. The clinical effectiveness of guided self-help versus waiting-list control in the management of anxiety and depression: a randomized controlled trial. Psychol Med 2005 Nov;35(11):1633-1643. [doi: 10.1017/S003329170500560X] [Medline: 16219121]

Abbreviations

CACE: Complier Average Causal Effect
CBT: cognitive behavioral therapy
HLP: healthy lifestyle program
iTAU: improved treatment as usual
MICE: Multiple Imputation with Chained Equation
MP: mindfulness program
NMW: national minimum wage
PANAS: Positive and Negative Affect Schedule
PAPP: positive affect promotion program
PHI: Pemberton Happiness Index
PHQ-9: Patient Health Questionnaire-9
SF-12: Short-Form Health Survey
VAS: visual analog scale

Edited by G Eysenbach; submitted 12.09.19; peer-reviewed by Y Leykin, J Lambert; comments to author 08.10.19; revised version received 03.12.19; accepted 04.02.20; published 05.06.20.

Please cite as:
Gili M, Castro A, García-Palacios A, Garcia-Campayo J, Mayoral-Cleries F, Botella C, Roca M, Barceló-Soler A, Hurtado MM, Navarro M, Villena A, Pérez-Ara MÁ, Riera-Serra P, Baños RM
Efficacy of Three Low-Intensity, Internet-Based Psychological Interventions for the Treatment of Depression in Primary Care: Randomized Controlled Trial
J Med Internet Res 2020;22(6):e15845
URL: https://www.jmir.org/2020/6/e15845
doi:10.2196/15845
PMID:32501276

Original Paper

Evaluation of Digital Technologies Tailored to Support Young People's Self-Management of Musculoskeletal Pain: Mixed Methods Study

Helen Slater[1], BAppSc, MAppSc, PhD, FACP; Jennifer N Stinson[2,3,4], RN-EC, PhD, CPNP; Joanne E Jordan[5], BA, BSc, MPH, PhD; Jason Chua[6], BSc; Ben Low[7], BA; Chitra Lalloo[3,4], BHSc, PhD; Quynh Pham[4,8], PhD; Joseph A Cafazzo[4,8,9], PEng, PhD; Andrew M Briggs[1], BSc (Hons), PhD, FACP

[1]School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University, Perth, Australia
[2]Lawrence S Bloomberg, Faculty of Nursing, University of Toronto, Toronto, ON, Canada
[3]Child Health Evaluative Sciences, Research Institute, The Hospital for Sick Children, Toronto, ON, Canada
[4]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[5]HealthSense (Australia) Pty Ltd, Melbourne, Australia
[6]Centre for Musculoskeletal Outcomes Research, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand
[7]Squawk Designs, Perth, Australia
[8]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
[9]Institute of Biomaterials and Biomedical Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Helen Slater, BAppSc, MAppSc, PhD, FACP
School of Physiotherapy and Exercise Science
Faculty of Health Sciences
Curtin University
GPO Box U1987
Perth, 6845
Australia
Phone: 61 892663099
Email: H.Slater@curtin.edu.au

Abstract

Background: Digital technologies connect young people with health services and resources that support their self-care. The lack of accessible, reliable digital resources tailored to young people with persistent musculoskeletal pain is a significant gap in the health services in Australia. Recognizing the intense resourcing required to develop and implement effective electronic health (eHealth) interventions, the adaptation of extant, proven digital technologies may improve access to pain care with cost and time efficiencies.

Objective: This study aimed to test the acceptability and need for adaptation of extant digital technologies, the pain*HEALTH* website and the *iCanCope with Pain* app, for use by young Australians with musculoskeletal pain.

Methods: A 3-phased, mixed methods evaluation was undertaken from May 2019 to August 2019 in Australia. Young people aged 15 to 25 years with musculoskeletal pain for >3 months were recruited. Phases were sequential: (1) phase 1, participant testing (3 groups, each of n=5) of co-designed website prototypes compared with a control website (pain*HEALTH*), with user tasks mapped to eHealth quality and engagement criteria; (2) phase 2, participants' week-long use of the *iCanCope with Pain* app with engagement data captured using a real-time analytic platform (daily check-ins for pain, interference, sleep, mood, physical activity, and energy levels; goal setting; and accessing resources); and (3) phase 3, semistructured interviews were conducted to gain insights into participants' experiences of using these digital technologies.

Results: Fifteen young people (12/15, 80% female; mean age 20.5 [SD 3.3] years; range 15-25 years) participated in all 3 phases. The phase 1 aggregated group data informed the recommendations used to guide 3 rapid cycles of prototype iteration. Adaptations included optimizing navigation, improving usability (functionality), and enhancing content to promote user engagement and acceptability. In phase 2, all participants checked in, with the highest frequency of full check-ins attributed to pain intensity (183/183, 100.0%), pain interference (175/183, 95.6%), and mood (152/183, 83.1%), respectively. Individual variability was evident for monitoring progress with the highest frequency of history views for pain intensity (51/183, 32.3%), followed by pain interference (24/183, 15.2%). For the goals set feature, 87% (13/15) of participants set a total of 42 goals covering 5 areas, most frequently for activity (35/42, 83%). For phase 3, metasynthesis of qualitative data highlighted that these digital tools were perceived as youth-focused and acceptable. A total of 4 metathemes emerged: (1) importance of user-centered design to leverage user engagement; (2) website design (features) promoting user acceptability and engagement; (3) app functionality supporting self-management; and (4) the role of wider promotion, health professional digital prescriptions, and strategies to ensure longer-term engagement.

Conclusions: Leveraging extant digital tools, with appropriate user-informed adaptations, can help to build capacity tailored to support young people's self-management of musculoskeletal pain.

(J Med Internet Res 2020;22(6):e18315) doi:10.2196/18315

KEYWORDS

musculoskeletal pain; mHealth; eHealth; self-management; adolescent; mobile phone; smartphone

Introduction

Background

Young people want access to health services that are tailored and responsive to their specific health needs, especially for those with chronic health conditions such as persistent musculoskeletal pain [1-3]. Digital technologies connect young people with health services and resources to support their self-care and promote positive health habits [4-6]. Furthermore, digital technologies are portable, customizable, and readily assimilate into young people's daily routines [5].

This capability to connect is especially important for young people with chronic health conditions, including pain, during the critical transition from childhood to young adulthood [1,7]. Persistent pain can impose a significant and enduring health and economic burden on young people and their communities [8,9]. Australian data suggest persistent pain rates for young people approach those of adults (ie, approximately 20%) [8,10], whereas international data indicate higher rates for musculoskeletal pain (eg, 37% for back pain) [11-13]. Young people often fall through gaps during this developmental transition, with many failing to seek care and adopting unhelpful habits that increase the risk of ongoing health and social issues [6]. In Australia, this risk is compounded by young people exiting pediatric health services around the age of 16 years and failing to seamlessly integrate into adult health services, despite transitional frameworks [1,14,15]. Digital technologies may help to drive engagement, enabling timely access and broader reach to the *right* health care [1,14,16,17]. Currently, the lack of accessible, reliable digital resources tailored to the needs and preferences of young Australians with persistent musculoskeletal pain remains a health services gap [14,18]. Recognizing the intense resourcing required to develop and implement effective electronic health (eHealth) interventions, adaptation of proven digital technologies may be a more cost- and time-efficient approach [19].

To accelerate the implementation of digital technologies into the care of young Australians with musculoskeletal pain, we established a transnational partnership to leverage digital technologies already developed and tested for pain care. Both platforms were co-designed using best practice recommendations for eHealth design and implementation [20]. First, in Australia, we developed a digital resource, pain*HEALTH*, to support improved musculoskeletal pain care for adults [21-27]. Co-designed with consumers, using a policy-into-practice approach [28], aligned to contemporary musculoskeletal models of care and strategic health frameworks (the chronic conditions framework [29] and the national pain strategy [15]), pain*HEALTH* supports consumers accessing the right care, at the right time, by the right team. Impact evaluation demonstrated that consumers/caregivers and health professionals perceived pain*HEALTH* as supporting holistic self-management and cocare of musculoskeletal pain [30]. Second, in Canada, an integrated smartphone and web-based self-management program, *iCanCope with Pain*, was developed [16,31]. This digital platform addresses the self-management needs of young people with persistent pain by improving access to contemporary tailored pain education; providing practical strategies to manage pain, psychological well-being, and sleep hygiene; encouraging physical activity; and providing peer social support [3,16]. The platform is undergoing cultural adaptation and usability testing evaluation in Norway [32].

Trial evaluation in Canada is also underway to assess implementation success and effectiveness (such as pain intensity, pain-related activity limitations, and health-related quality of life) outcomes (clinical trial number: NCT02601755).

Objective

Therefore, to avoid research waste and technology duplication, this study aimed to test the acceptability and need for adaptation of these two extant digital technologies for young Australians with musculoskeletal pain. The specific aims were as follows:

1. To test the acceptability and usability of a prototype derivative of pain*HEALTH* for young people with musculoskeletal pain
2. To test the acceptability and usability of the *iCanCope with Pain* app [16] for young people with musculoskeletal pain.

Methods

Study Design

This 3-phased, mixed methods evaluation was undertaken in Australia between May 2019 and August 2019. Phases were sequential, with phase 1 involving participant (user) testing of website prototypes (research aim 1). In phase 2, participants utilized the *iCanCope with Pain* app (hereafter *app*) over a 1-week engagement period (research aim 2). Finally, in phase 3, semistructured interviews were conducted to provide insights into the participants' experiences of using these digital technologies (Figure 1). The study had institutional ethics approval, adhered to the Declaration of Helsinki, and aligned with reporting recommendations from the consolidated criteria for reporting qualitative studies: 32-item criteria (Multimedia Appendix 1) [33].

Figure 1. A 3-phased approach to the user testing was adopted, with all participants undertaking user testing of website prototypes (phase 1), pain app user testing (phase 2), and semistructured interviews on their experiences of using these digital technologies (phase 3).

Eligibility Criteria

Young people aged 15 to 25 years with musculoskeletal pain for >3 months and currently living in Australia were eligible to participate. Musculoskeletal pain included recurrent or persistent pain; however, pain at the time of enrollment was not a prerequisite for participation. These eligibility criteria were aligned to our previous research [1] and applied across all 3 study phases.

Recruitment and Sampling

Participants were recruited from community sources, including arthritis consumer organizations, private health care practices, youth mental health services, via social media, electronic newsletters, flyers, and emails. Initial eligibility screening was undertaken using a web-based survey platform Qualtrics. The survey instrument was informed by our previous research on young people with musculoskeletal pain [1,16], with minor iterations, and was piloted by members of the research team (HS, AMB, J Chua, and JNS) before deployment to the field. No researcher had an existing relationship with any participant.

Consent

Young people meeting all 3 inclusion criteria were requested to provide consent directly through the Qualtrics web platform. Participants aged 15 to 17 years were advised to discuss consent with their guardian/parent before consenting to participate and were also invited to have a guardian/parent present for any, or all, phases of the study.

Protocol

The participants meeting the inclusion criteria and providing consent completed an initial web-based survey to capture demographics and clinical characteristics, including area(s) of their pain, pain duration, and pain diagnosis. The Örebro Musculoskeletal Pain Screening Questionnaire-Short Form (ÖMPSQ-SF) [34] was included as a measure of pain-related disability. The ÖMPSQ-SF has been validated for use in primary care [35] and used for previous clinical pain research in this age group [1]. Items are scored 0 to 10; 0 refers to the absence of impairment and 10 to severe impairment. A total of 3 items are reversed in order for all the questions to be oriented in the same direction. Total scores range between 1 and 100, with a score >50 indicating higher estimated risk for future (work) disability. Given the age range for our cohort, we modified the wording for items where *work* was mentioned to include *work/study*. Once participants completed the web-based survey, they were contacted to schedule a time for phase 1 of the user testing.

Participants were allocated sequentially to 3 testing groups (each comprising 5 participants). Groups were run in series, with group allocations remaining consistent across all 3 phases of testing (as all participants consented to complete all 3 phases). The purpose of serial testing in phase 1 was to enable iterative cycles of formative feedback used to guide prototype refinements before the subsequent testing group, an approach consistent with best practice recommendations for digital user testing [16,20]. The period of testing across all 3 phases for a group ranged from 11 to 26 days. At the completion of phase 3, all participants received an AUD $100 (US $65) gift voucher.

Phase 1. painHEALTH and Prototype Website User Testing

For this phase of user testing, the existing pain*HEALTH* platform was used as a *control* website. This allowed users to evaluate 2 newly developed website prototypes against a fully functional comparator. As pain*HEALTH* was co-designed with adults, rather than younger Australians, we did not expect participants to easily relate to this website, but rather to consider features and content that might be helpful on a youth-focused platform. The prototypes were informed from our previous research using insights from young people with musculoskeletal pain about their needs and preferences for digital tools [1]. The specific tasks are outlined later in the paper.

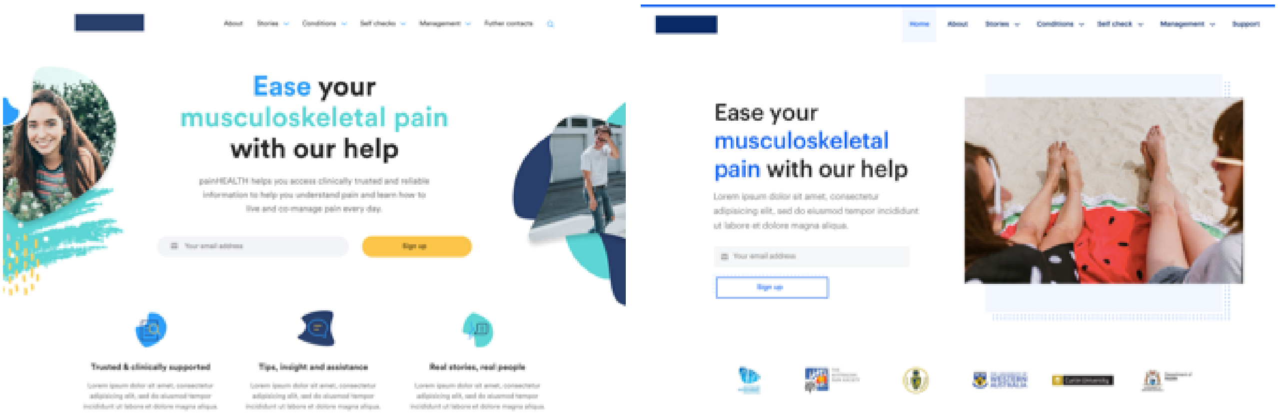
We implemented *Lookback* as the test platform, using the *LookBack Live* option. *LookBack Live* allows users and researchers to engage and communicate together in real time with a website (in this case, pain*HEALTH* and website prototypes) via a shared screen, with reactions, comments, behaviors, and interactions captured via downloadable audiovisual recording. This platform provided a good *fit for purpose*, enabling flexible scheduling, remote user access, and participants use of their own computers or mobile devices without the need for additional cables or cameras. User insights are denoted by a researcher's use of *time stamps*, with additional capacity for taking technical notes within the test platform that can later be extracted in text and recorded form. A rapid cycle of prototype iteration was undertaken at each group's completion of user testing.

Before implementing the user testing, the LookBack Live platform and user tasks were piloted by members of the research team (HS, AMB, JNS, and CL). Minor amendments were made to the user tasks to improve clarity regarding tasks. Participants were scheduled to a time slot, emailed a unique LookBack Live invitation link, and provided with instructions for the session. Participants were informed that the user testing was about the prototypes (and not themselves) and they "can't do or say anything wrong" and to "think aloud" as much as possible. The total mean duration of the web-based prototype testing was between 60 and 90 min.

Participants were tasked with evaluation in the following order:

- The pain*HEALTH* website: This website was used as a control site to help orient participants toward the look and feel, functionality, and content of a contemporary, consumer-focused, Australian musculoskeletal pain website [30]. Users were guided through 4 tasks covering interactive content (pain conditions and stories, self-checks, and pain management) on musculoskeletal pain care.
- Two newly developed prototype websites (Figure 2): Prototype 1 was characterized by soft round shapes with a palette of purple/yellow/pink, and prototype 2 featured a geometric design with a blue/black palette). Content for the prototypes was adapted from the pain*HEALTH* website for the evaluation tasks. The prototype websites were powered by InVision (digital product design).

Figure 2. Screenshots of the 2 prototype websites: to the left, the round, soft version is shown, and to the right, the geometric version.



Participants completed the same 4 tasks across each of the 3 websites (Multimedia Appendix 2). Each test task was contextualized by: (1) a background statement about the nature of the task, (2) instructions to guide the specific task required, (3) specific questions about each task to gauge feedback, and (4) any additional feedback. The 4 test tasks were as follows

1. Find and watch *Daniel's pain story*: This video/text content captured a young person's narrative around their experiences with low back pain.
2. Find and complete the '*Örebro Musculoskeletal Pain Self-Check*': This self-check questionnaire is designed for use across musculoskeletal pain conditions to predict the risk of higher estimated disability [34], and users can print out their score to share with health professionals.
3. Find and read the *Making Sense of Pain* management module: This content is focused on reasons why pain can persist beyond normal tissue healing times and the need for tailored holistic and integrated pain care.
4. Find *Further assistance*: This page provides a list of West Australian pain services and links to a cross-discipline health professional listing managed by the Australian Pain Society.

These test tasks were designed to reflect user-centered eHealth design principles [20] and criteria that predict real-world user engagement [20,36,37], specifically:

- Navigation
- Usability (functionality and ease of utilization)
- User engagement (content presentation, interactive, not irritating, targeted/tailored/personalized, captivating, and relatability)
- Content (appropriate level of literacy, credible, clear, and concise)
- Acceptability (design look and feel, meets the expectations of young people with musculoskeletal pain, motivating, and likability).

Engagement criteria were intentionally mapped across the task tests, with some overlapping of criteria (Table 1) to enable meaningful interpretation of outcomes and inform prototype iteration.

Table 1. Four test tasks were mapped, as per ticks, to reflect specific electronic health evaluation criteria.

Design and quality criteria	Test tasks			
	Daniel's story	Self-check[a]	Making sense of pain	Further support
Navigation	✓	N/A[b]	N/A	✓
Usability	N/A	✓	N/A	✓
User engagement	✓	✓	✓	✓
Content	✓	✓	✓	✓
Acceptability	✓	✓	✓	✓

[a] Self-check refers to validated self-report questionnaires.

[b] N/A: not applicable.

Phase 2. iCanCope With Pain App User Testing

For phase 2, we adapted the framework developed by Stinson et al [16] for guiding young people's use of the *iCanCope with Pain* integrated web- and smartphone-based app [16,31]. The specific app features [16] tested included:

- Symptom(s) check-in
- Structured goal setting to improve pain and function
- An interactive toolbox for pain coping strategies
- Access to papers to support users understand and manage their pain, mood, sleep, activity, and social function.

Participants were contacted by email within 2 days of completion of phase 1 and provided with the following standardized information (Multimedia Appendix 3):

- An overview of the *iCanCope with Pain* app (why it was developed, how it is currently being used, and by whom) and a link to the *iCanCope with Pain* website
- Features of the app used to support the self-management of pain (tracking and monitoring function; tracking your sleep, mood, physical activity, and energy levels; setting realistic goals; coping tools; and resources to support care)
- Instructions on how to access the app (via the app store) and sign in using their unique independently generated user name and password.

Instructions also outlined 4 key tasks for participants to complete:

1. Setting up their individual user profile (pain areas, triggers, pain intensity, interference, mood, sleep, physical activity, and energy)
2. Setting up their individualized goals aimed at improving their functioning in 5 main areas (physical activity, sleep, social, mood, and energy) and monitoring their progress against these goals
3. Symptom tracking over the 7 days to self-monitor progress (including pain, interference, mood, sleep, physical activity, and energy) and captured in the form of daily *check-in* reports
4. Accessing and exploring the library and the community resources designed to support their pain management (topics including nutrition, mood, sleep, work and study, social activities, self-worth, coping with pain, dealing with setbacks, talking to friends, employers, and teachers about pain).

Once set up and having completed these initial tasks, participants were encouraged to use the app daily to self-track and monitor their progress over the following 7-day period, while also using the app as they wished. Data extraction was performed by a team member (QP) blinded to phase 1 outcomes.

Phase 3: Semistructured Interviews

In the final phase, individual in-depth interviews were undertaken with participants about their experiences of using these 2 digital tools (website and app). The interview schedule was developed by a multidisciplinary research team (including clinical researchers with experience of musculoskeletal pain [HS, JNS, AMB]) and informed by prior user testing of these digital platforms [16,30] and our research on young people experiencing musculoskeletal pain [1,14] (Multimedia Appendix 4). Before the interview, participants were contacted (JEJ) to outline the interview process and schedule a time for the interview, in a quiet location chosen by the participant. All 15 interviews were conducted via teleconferencing by a senior qualitative researcher (JEJ) with experience in interviewing young people with musculoskeletal pain and who was independent of phase 1 and 2. No participant requested a support person during the interview. Interviews were audio recorded (duration ranged between 14 and 36 min, mean duration 21 min) and transcribed verbatim. Verbatim transcripts were checked for accuracy.

Data Analysis

Phase 1

User insights and feedback were extracted at the completion of each group's cycle of testing. These text-based group data were recorded in an Excel spreadsheet, aggregated (BL), and mapped against the established eHealth criteria (HS; Table 1). These data were discussed by members of the team (HS and BL), and a consensus reached about what iterations should be implemented. Rapid cycles of prototype iteration were then undertaken and implemented before the next cycle of group testing. Therefore, 3 cycles of prototype iteration were undertaken and implemented (between groups 1 and 2, between groups 2 and 3, and 1 final cycle at the completion of group 3 testing).

Phase 2

Participant engagement over the 7 days of testing was captured and evaluated through an Analytics Platform to Evaluate Effective Engagement (APEEE) with digital health interventions [38]. APEEE is a dynamic, real-time analytic platform that captures and characterizes user app engagement. Outcomes from APEEE included the following:

- Total number of full check-ins (completion of all relevant domains), time taken to check-in (min/second), check-in by domain (pain activity, mood, physical activity, sleep, and energy). Note that sleep can only be logged once per day.
- History views (number of times each symptom domain was reviewed: pain, interference, activity, mood, sleep, and energy). Note that pain intensity is the default history domain, with users having to toggle to view other domains.
- Goals set (number and domain: activity, sleep, energy, mood, social).
- Library articles accessed (number and type).

Phase 3

The 15 verbatim transcripts were analyzed (coded) by 1 researcher (JEJ) in 3 sequential stages (ie, 3 sets of 5 transcripts) using inductive and deductive approaches. In the first stage, 5 transcripts were analyzed using a general inductive analytical approach, where codes were directly derived from the data or from *the ground up*, without starting from a prior theoretical understanding of the issue being explored. For the second stage, the coding framework inductively derived from the first five 5 transcripts was then utilized to deductively code the second set of 5 transcripts. When a new topic emerged from the data, a corresponding code was developed inductively. In the third stage, to verify data redundancy, the revised coding framework was applied to the final set of 5 transcripts. No new codes were added to the framework in this final stage of the analysis. By utilizing a combined inductive and deductive analytical approach, which involved a constant comparison of data over a period until no new topics emerged (data redundancy), a comprehensive profile of participant preferences and experiences was developed. Following the finalization of the coding framework, codes were reviewed, and key themes and corresponding subthemes were developed through an iterative process of grouping codes into concepts, reviewing transcripts, and refining themes. The interview schedule was intentionally divided into 4 parts: (1) general background/context (questions 1 and 2); (2) perceptions of the pain*HEALTH* website functionality (question 3); (3) perceptions of the *iCanCopewith Pain* app functionality (question 4); and (4) perceptions related to the acceptability, use, and implementation of the 2 digital tools generally (questions 5-13). This enabled the analysis of transcripts to be undertaken for each part of the schedule as mutually exclusive categories. Key themes and subthemes aligning to each part of the interview schedule were then reviewed (HS, AMB, and JEJ) and grouped under overarching (meta) themes [39]. One-third of the interview transcripts (n=5) were independently analyzed (JNS and CL) to confirm themes identified, and where necessary, refined to reach consensus, and confirm the construct validity. Data were structurally organized to present themes in a logical explanatory scheme [40].

We evaluated the extent to which our inductively derived data mapped to an evidence-based eHealth roadmap (the latter, extensively reviewed elsewhere) [20]. The roadmap takes a holistic approach to the development of eHealth technologies, integrating persuasive health technology theories with business modeling (efficient, effective, and sustainable) to improve the uptake and impact of eHealth technologies in practice. The roadmap is agnostic to digital platforms, is inherently fluid, interactive, and allows for a cyclical approach to the co-design (acceptability, functionality, and usability) and implementation of digital technologies in real-world settings. The roadmap provided a good *fit for purpose*, the plurality allowing for analysis of content relevant across both digital platforms (specifically, contextual inquiry, user requirements, value specification, co-design, and operationalization, referred to here as implementation).

Results

Demographic and Clinical Characteristics

A total of 20 potential participants met the inclusion criteria and consented to participate (Table 2). Of these 20, 15 (75%) participated and completed all 3 phases of the study, and 5 recruits did not start the study for various reasons (2 unwell, 2 for personal reasons, and 1 did not respond to email contacts or phone follow-up). The majority of these 15 participants were female (12/15, 80%) and resided in Western Australia (13/15, 87%). The mean age of these participants was 20.5 (SD 3.3) years.

For those participants with a confirmed medical diagnosis (10/15, 67%), musculoskeletal conditions included Ehlers-Danlos Syndrome (n=2), fibromyalgia (n=4), endometriosis (n=1; comorbid with back pain), rheumatoid arthritis (n=1), scoliosis (n=1), sacroiliitis (n=1), hip labral tear (n=1), and low back pain with nerve-related leg pain (n=1). For those reporting no confirmed diagnosis, conditions primarily related to low back/neck pain or upper/lower limb muscle or joint pain.

Table 2. Demographic and clinical pain characteristics of consenting participants (N=20).

Demographic/clinical pain characteristics	Consenting and participating (n=15)	Consenting and not participating (n=5)
Age (years)		
Mean (SD)	20.5 (3.3)	22.6 (3.1)
Range	15-25	18-25
Gender (female), n (%)	12 (80)	4 (80)
Urban/rural, n (%)	13 (86)	1 (20)
English as a first language, n (%)	14 (93)	5 (100)
Highest current level of education completed, n (%)		
University	4 (26)	1 (20)
TAFE[a]	1 (6)	2 (40)
Year 12 (tertiary entrance)[b]	6 (40)	1 (20)
Year 12 (other)	2 (13)	1 (20)
Less than 3 year secondary	2 (13)	0 (0)
Currently at, n (%)		
School	4 (26)	0 (0)
University or TAFE	8 (53)	1 (20)
Unemployed	0 (0)	2 (40)
Employed (volunteer or paid work)	3 (20)	2 (40)
Pain		
Diagnosis from health professional (yes), n (%)	10 (66)	4 (80)
Duration of pain (years)		
Mean (SD)	6 (6)	8 (8)
Range	0.3-22	2.5-18
ÖMPSQ-SF[c]		
Mean (SD)	47 (14)	62 (6)
Range	27-74	52-66
Area(s) of pain[d], n (%)		
Neck pain	6 (40)	3 (60)
Mid back	7 (46)	3 (60)
Low back	8 (53)	4 (80)
Hips	5 (33)	2 (40)
Knees	3 (20)	2 (40)
Ankles	4 (26)	0 (0)
Shoulders	4 (26)	1 (20)
Elbows	1 (6)	1 (20)
Wrists/hands	11 (73)	1 (20)
All over pain (muscles and joints)	6 (40)	2 (40)
Other pain[e]	10 (66)	3 (60)

[a] TAFE: Technical and Further Education Institutions.

[b] Pathway for university entrance.

[c] ÖMPSQ-SF: Örebro Musculoskeletal Pain Screening Questionnaire-Short Form, possible score 1 to 100.

[d] Total count may be greater than the number of participants as more than one area of pain could be nominated.

[e] Areas of pain nominated in free text included abdominal pain (n=3), coccygeal pain (n=1), migraine (n=3), gastrointestinal issues (n=2), dysmenorrhea

(n=1), and nerve pain (n=1).

Phase 1: pain*HEALTH* and Prototype Website User Testing Outcomes

Outcomes from user testing of websites provided rich insights across the 3 groups. Outcomes are summarized sequentially by groups (1, 2, and 3) with recommendations derived to inform each cycle of prototype iteration, shown in Textboxes 1 to 3, respectively. A comprehensive tabulation of these data mapped by group to design criteria (navigation, usability, user engagement, content, and acceptability) with supporting user insights and quotes is provided in Multimedia Appendix 5.

For group 1, outcomes are summarized in Textbox 1. Participants leaned more strongly toward preferring the *round and bright* prototype variation. Compared with the control site, participants liked the use of dropdowns (easier navigation) and self-check functionality enhancements (better user experience for quiz completion). The *Further Support* menu did not resonate well with participant expectations, as this suggested *technical support* for some. Simplifying language and the use of a glossary were perceived as enhancements that would improve user engagement. On the basis of these outcomes and discussions with members of the team (AMB, JNS, BL, and HS), a consensus decision was reached to proceed with the *round and bright* prototype only. Collectively, these data informed the first rapid cycle of prototype iteration, implemented before group 2 testing.

Textbox 1. Recommendations for prototype iteration based on insights from group 1 participants.

- Navigation: add drop-down menus throughout the prototype website to facilitate navigation and improve user engagement.
- Usability: *self-check* improvements included providing an explanation about what the self-check results mean to help reduce anxiety and fear about the results and the addition of explanatory text about who to see (health professional) and why.
- Content and user engagement: simplifying text, chunking text, and use of bold quotes to highlight key messages.
- Acceptability: *Further Assistance* changed to *Further Help* to capture both the listing of resources, services, and links to health professionals.

Group 2 insights indicated overall positive responses to the colorful prototype iterations and informed further recommendations for iteration (Textbox 2; Multimedia Appendix 5). Participant insights highlighted the ease of navigation, usability (functionality), engagement, content, and acceptability compared with the pain*HEALTH* website. Recommendations to inform the second cycle of prototype iteration were implemented before group 3 testing.

Textbox 2. Recommendations for prototype iteration based on insights from group 2 participants.

- Navigation: enhance the navigation bar to ensure it is *sticky* on every page (ie, the navigation bar follows the user as they move up and down a page); implement hyperlinks to open external websites in a new tab; improve identification of external hyperlinks versus internal; add global navigation to the self-check, with the removal of progression steps; inclusion of a progress bar/percentage; remove lifestyle image on self-check start page; and use a smaller icon to reduce the need for users to scroll to start the self-check quiz.
- Usability: for the self-check, revise the text to better explain what the self-check is, why a user would complete it, and how they can use this information.
- User engagement: introduce categories within the management modules to better direct users to relevant content and further optimize and enhance user engagement and acceptability.
- Content: review website content and optimize for a younger reading audience (the appropriate level of health literacy) with even more *chunked content* and amend *Further Contacts* further, adding 3 tabs to be meet user expectation about related page content and to reduce the volume of contacts listed on the page into clear groupings to be easily scannable.
- Acceptability: color palette optimization to improve the relatability for the targeted age range using a softer palette of welcoming blues and engaging yellows that still *feels* young; drop the pinks and purples to ensure accessibility for color-blind individuals.

Group 3 participants indicated a positive overall response to the prototype compared with the control site (Textbox 3; Multimedia Appendix 5), with all indicating that the color palette was engaging and fun and nicely targeted toward a younger demographic. Feedback highlighted an improved user experience in navigation, with no further issues reported. Self-check functionality enhancements were well received. One participant raised the possibility of implementing screen reader functionality to help improve readability (when required). Compared with the control site, the use of chunked information and larger font size made for easier readability. Compared with the control website, this prototype was perceived by group 3 participants as providing a good example of a holistic, integrated approach to young people’s pain, making it easy to find, research and understand their pain. To further support and explain content, the use of more illustrations/videos within the management modules was proposed. Overall, participants found the website acceptable, reporting the look and feel as engaging, fun, relatable, and appropriate to their demographic. Collectively, group 3 insights guided the final prototype iteration recommendations.

Textbox 3. Final recommendations for prototype iteration based on insights from group 3 participants.

- Usability: consider screen reader functionality to support users or consider extracting the audio from website videos, loading these as audio files. For longer form content, capturing readings of content, to then create downloadable audio files.
- Content: the use of a sidebar section that can act as an *anchor link* that is sticky and follows the user up/down the screen and scrolls their viewpoint to the contact area to further assist finding relevant contacts.
- User engagement: embed images that are resonant of the specific pain management module the content is about (eg, an image or icon of an activity or walking for *Movement with Pain*) to help differentiate the management module content and better assist, engage, motivate, and relate to the user.

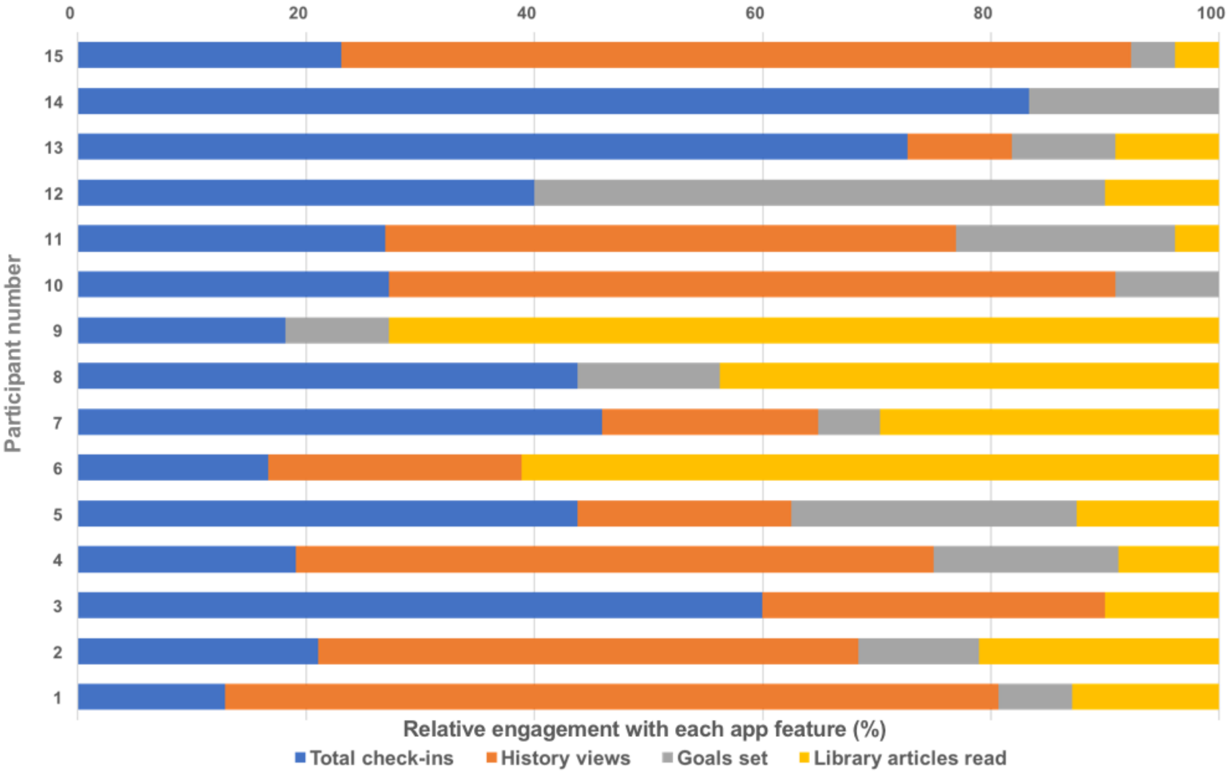
Phase 2 Outcomes: *iCanCope with Pain* App

Participants' engagement data for the use of the *iCanCope with Pain* app are summarized in Table 3 and Figure 3. The group mean (SD) and median frequencies for specific app features were: check-ins, 7.3 (4.0), 7.0; history views, 10.5 (12.8), 4.0; goals set, 2.8 (2.8), 2.0; and articles accessed, 4.2 (3.9), 2.0. Over the week of app use, all participants checked in, with 183 check-ins initiated. Of these, 59.6% (109/183) were full check-ins, meaning completion of all 5 domains (sleep, mood, physical activity, and energy levels). The highest frequency of full check-ins was attributed to pain intensity (183/183, 100.0%), followed by pain interference (175/183, 95.6%), and then mood (152/183, 83.1%). Attrition for check-ins was more common for domains covering physical activity (120/183, 65.6%) and energy (109/183, 59.6%). Note that the sleep domain is only asked once per day regardless of the number of times participants checked in each day (therefore, no denominator provided), which explains the lower number relative to other domains (n=72).

Table 3. Individual participant engagement data for the use of the *iCanCope with Pain* app over 7 days.

Participant ID	Full check-ins, n	History views, n	Goals set, n	Library articles accessed, n
1	4	21	2	4
2	8	18	4	8
3	6	3	0	1
4	13	38	11	6
5	7	3	4	2
6	3	4	0	11
7	17	7	2	11
8	7	0	2	7
9	2	0	1	8
10	6	14	2	0
11	7	13	5	1
12	4	0	5	1
13	8	1	1	1
14	5	0	1	0
15	12	36	2	2

Figure 3. Individual-level data are shown for participants 1 to 15 (vertical axis). Relative user engagement (horizontal axis, proportional frequencies, and ranges for each domain) is presented across 4 key domains: (1) total (full) check-ins (blue), range 2-17; (2) history views (bright orange), range 0-38; (3) goals set (gray), 0-11; and (4) library articles accessed (light orange), range 0-11. Note that it is the variable relative engagement of each individual with the features of the app.



A total of 73.3% (11/15) participants completed a total of 158 history views over the week of app use (ie, monitoring their progress over time), with a variable frequency of views for each domain. The highest frequency of domain history views related to pain intensity (51/158, 32.3%), followed by pain interference (24/158, 15.2%), activity (22/158, 13.9%), mood (21/158, 13.3%), sleep (19/158, 12.0%), and energy (18/158, 11.4%).

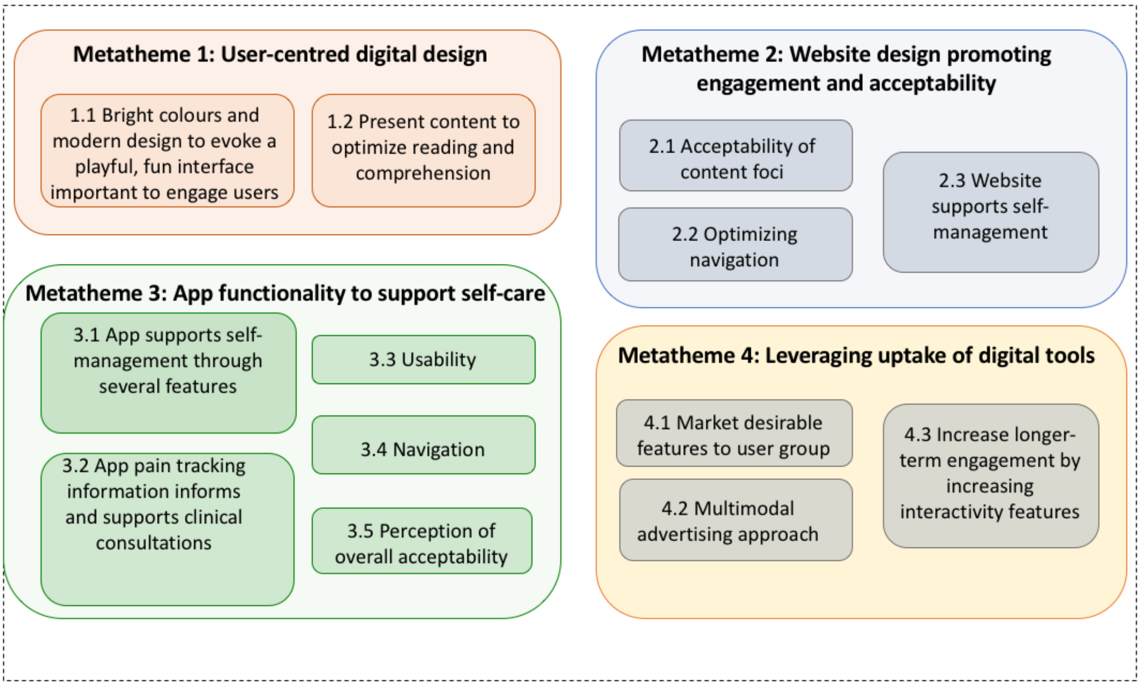
For the *goals set* feature, 86.7% (13/15) of participants set a total of 42 goals covering 5 areas. The most frequently set goals were for activity (35/42, 83.3%), with fewer for sleep (4/42, 9.5%), energy, mood, and social activity (1/42 for each, 2.4%). Over the week of app use, 63 articles were accessed through the app library by 11 (73.3%) participants. Articles focused on the following content areas (listed most to least frequently accessed; note *n* may be greater than participant numbers, as more than one article could be categorized within 1 content area):

- Making sense of pain (16/63, 25%)
- Sleep hygiene (14/63, 22%)
- Behavioral approaches to pain (coping and mood management; 10/63,16%)
- Activity planning, pacing, and goal setting (8/63, 13%)
- Communicating pain (4/63, 6%)
- Exercising with pain (4/63, 6%)
- Managing fatigue with pain (2/63, 3%)
- Condition-specific care (neuropathic pain; 1/63, 2%)
- Motivation and adherence (1/63, 2%)
- Nutrition (healthy eating habits; 1/63, 2%).

Phase 3: Qualitative Interview Findings

Findings are presented according to a hierarchical structure aligned to the qualitative data analysis methods, with 4 overarching metathemes identified, supported by themes and subthemes (Figure 4). An overarching summative statement for each metatheme is provided in the text later in the paper. Further details for metathemes, including specific themes and subthemes with supporting quotes, are summarized in Textboxes 4 to 7, respectively. A comprehensive tabulation of all findings is provided in Multimedia Appendix 6.

Figure 4. Graphic summary of metathemes and themes derived from qualitative interviews. Metathemes were as follows: user-centered digital design (orange), website design promoting engagement and acceptability (blue), app functionality to support self-care (green), and leveraging uptake of digital tools (yellow).



Metatheme 1: User-Centered Design Features Resonating With Young People's Needs

User-centered design features perceived by participants across both digital platforms as reflecting specific youth-focused digital tools included vibrant color palettes, playful nonlinear shapes, and engaging interfaces that motivated them to interact with the digital tools. Participants described content across both platforms in a manner that was relatable and engaging for young people, with the use of simple language, a variety of formats to deliver information, and chunking of text, rated as preferred features (Textbox 4).

Textbox 4. Metatheme 1: user-centered digital design.

Bright colors and modern design that evoke a playful and fun interface are important features of digital tools (website and app) to engage users

- Vibrant colors associated with being youth-focused

  "I felt like the colours were good. I felt like the design was youth-focused, but not necessarily exclusively youth-focused, if you know what I mean? If a 50-something year old were to look at it [the app], they wouldn't just be like, 'Oh, this is dodgy.'" [Participant number 11; app]

- Vibrant colors associated with eliciting positive emotions

  "I really enjoyed the bright colour scheme with the blues and the purple and the pinks. I thought it was really calming and really interesting, as well, for a young audience." [Participant number 9; website prototype 1]

- Different shapes, curved edges, lots of images, and illustrations are modern design features that appeal to users

  "I really enjoyed how the layout was all the bubbles instead of being just straight text or in boxes, it seemed really interesting and fun." [Participant number 9; app and website]

Present content to optimize reading and comprehension

- Use of different formats to deliver information (eg, video, audio, text, illustrations)

  "I found it all very easy to read. It was nice big font with a good balance between images to break up the text, as well as videos and stuff like that as well." [Participant number 3; prototype websites]

- Careful attention to font—the size, type, and color all impact readability (particularly if user is feeling unwell/fatigued)

  "But the writing itself is a decent size, so for anyone that does have a bit more trouble with that, like if they do have a headache, the fact that it's a bit bigger is better as well." [Participant number 1; app and websites]

- Using positive, nontechnical language, and less formal wording to enhance engagement

  > *"I guess the thing that struck me the most, I guess, in a good way was how it was sort of almost unprofessional, like it was just asking these questions with almost like childish cartoony images."* [Participant number 3; app]
  >
  > *"It's not 'for kids,' there's no being spoken down to at any point with this site."* [Participant number 10; website]

- Short and concise information with links to additional information/external resources

  > *"The shortness of them, the conciseness is pretty helpful as well, so I can just sit down for 5 minutes, read it and then take it on, see if it works and if it doesn't work, find another one, like, find a method to suit me better."* [Participant number 14; app and prototype website]

Metatheme 2: Website Design Promoting User Acceptability and Engagement

Participants identified features of the website prototypes that aligned with their preferences for acceptability, ease of navigation, and usability of digital tools. Acceptability here referred mainly to the information being relevant to participants (Textbox 5). Where content was too text-heavy (ie, pain*HEALTH* website), participants suggested modifying density by *chunking information*, using *callouts*, and greater use of supporting images/interactive features (such as video content). Participants liked the use of real-world patient stories on the websites, which resonated with their individual experiences of pain, reduced feelings of isolation, and showed how others had learned to cope with their pain. Navigation features that motivated participants to engage with the website included drop-down menus, clear website layout, simple design, use of a search bar, and minimal clicks to get to relevant content. Participants expressed enthusiasm for a tailored website specific to the needs of young people with pain. Features including meaningful pain management information, practical tips, real-world pain stories, and insights into condition-specific care were positively perceived to support their pain self-management.

Textbox 5. Metatheme 2: website co-design promoting user acceptability and engagement.

Acceptability of content foci

- Information matches user needs

  > *"It's got plenty of information on different types of pain, things like sleep, which is really helpful as well 'cos that's a big part of it, and also information on who you could speak to if you're having further issues, which is really nice as well because there's not a lot of that information elsewhere."* [Participant number 11; pain*HEALTH*]
  >
  > *"Also, I found that the stories themselves are pretty handy… I think it shows me that I'm not the only one doing it, but they're also demonstrated methods of coping with it within their everyday life. That was the key bit."* [Participant number 14; pain*HEALTH*]

- Information pages too text-heavy and need to be broken up with the use of images, diagrams, and callout text features

  > *"I felt further work could be done to improve things like having a lot more figures and images in the text, so like the articles or explanations, having a lot more diagrams that are explaining some of the science."* [Participant number 22; pain*HEALTH*]

- More information wanted on services available and how to access help

  > *"...I think the Help section, where it gave the list of the hospitals and healthcare centres, that needs to be a bit more refined so that it's easier to see what you were looking for. I'd like to see a website, like a hospital website and then further contact details on that website or instructions on how to get to the details on the service and what they can do."* [Participant number 6; website prototype 1]

Optimizing navigation

- Drop-down menu feature widely liked and assisted easy navigation

  > *"Everything is easy to find because it's labelled very nicely at the top there for you and it's all dropdown, so it's really easy to go into what you're looking for."* [Participant number 1; website prototype 1 and 2, version 1]

- Website layout design clear and easy to use

  > *"It's very clearly laid out and really easy to navigate without being, I don't know, patronising. I've been on patronising information sites before."* [Participant number 10; website prototype 1]

- Search bar considered an important feature that would assist looking for specific information

  > *"The search bar would make it easier, but it was pretty straightforward and I was able to find what I was looking for."* [Participant number 4; website prototype 1 and 2, version 1]

- Preference for minimal clicks to access information

  > *"Probably an update to the home page so when you click on the menu bar you don't go to another page that's got all the links there, you actually just go directly to the page that you're looking for from the menu bar."* [Participant number 14; website prototype 1]

Website supports self-management

> *"It's amazing to think that things like this are coming out because when you first get diagnosed or whatever there really aren't places like this. So it's exciting to see that people are starting to think about this. It's great."* [Participant number 12; website prototype 1]

Metatheme 3: App Functionality Supporting Pain Self-Management

Participants identified multiple app functionalities that they perceived supported active self-management of their pain. Notably, the daily check-in was seen as encouraging self-reflection and identification of broader factors that may impact their pain, whereas pain tracking information was useful in supporting their pain care through positive behavioral change and guiding clinical pain encounters (Textbox 6).

Usability features that motivated participants to engage with the app included the daily check-in (easy to use), content prompts based on input from check-ins (eg, mood management), and being able to connect with young peers. Participants identified several app functionalities that could enhance daily check-ins, including:

- Ability to set a flexible check-in time that suited their daily routines (versus the standard midday check-in)
- Functionality that allows for mapping output from daily activities and personal diary functionality for reflection/reminders
- Retrospective diary entry capability (in case of missing a day of entry)
- Capacity for entries in different pain areas (ie, pain intensity)
- Alternate mechanisms for ratings at check-in (versus spin a dial)
- Reminder notifications remain active until you have completed a task.

Enhanced functionality to improve self-management included enabling more specific alignment of user goals with measurable outcomes (ie, what goal did you set? did you achieve the goal?) and a metric to map outcome (eg, how much exercise did you achieve?) tailored to the individual. Some participants suggested improved functionality to enable printing and emailing of pain tracking information would be beneficial, as would the ability to personalize an avatar. Participants were largely positive about app navigation functionality, with features indicating strong usability highlighted as intuitive design, ease of use, and accessible content. A majority of participants (10/15, 67%) disliked the app icon *Copey*, indicating that it was more appropriate for younger users. Overall, the app was acceptable to all participants and perceived as a valuable digital tool to promote self-management of pain, with 24-hour access via their smartphones, enabling a flexible fit within their daily routines.

Textbox 6. Metatheme 3: app functionality to support young people's pain self-management.

***iCanCope with Pain* app supports self-management through several features**

- Daily check-in encourages self-reflection and identification of broader factors that may impact on pain

  > *"So I think it was good for bringing to light some of those things, because normally you wouldn't typically finish your day and be like, 'I was in a lot of pain today. Why was I in a lot of pain?' and this helps to start those, I suppose."* [Participant number 3]

- Pain tracking supports positive pain care behaviors (monitoring, self-reflection, goal setting, coping, and progress over time)

  > *"It was interesting to see how much sleep I got related to my pain and my mood etc... It's like, maybe I should sleep more."* [Participant number 6]

- Encourages setting of goals and monitoring progress

> *"That was awesome. That part was really good because you can track personally what your goals are, when you've completed them, whether they're short-term goals or long-term goals."* [Participant number 13]

Utilize pain tracking information to inform and support clinical consultations

> *"I think it could be used for both, like for just monitoring it yourself. If health professionals need to see how it's changing, that would be a good tool for them to use."* [Participant number 4]

Usability

- Daily check-in feature engaging and easy to use

> *"I just like how easy it is. It doesn't take very long to check-in and it's very easy to see how it's changing and what's impacting it."* [Participant number 4]

- Information interactivity features appealing

> *"I think after you enter it for three days it starts giving you some recommendations and patterns, which was nice. It gives you feedback and I think it was giving some suggestions of articles, which was good, then you click on them and start reading."* [Participant number 7]

- Copey, a divisive character—mixed perceptions regarding the acceptability of the app icon for the target user group

> *"I quite liked it, but I think it might be a bit too childish for my demographic. The icon is iconic, but I think that's what is making it childish for the younger group."* [Participant number 6]
>
> *"I think it is. I think it's a little childish, but I think that young adults are returning to enjoying more childish things...maybe having the icons be a little bit more adult."* [Participant number 9]

- Multifaceted improvements for daily check-in feature

> *"Yeah, it doesn't have anything that specifies [what time] I'm checking in and also it seems to let me check in multiple times a day."* [Participant number 10]

- Greater functionality (design features) to support end user self-management

> *"Also with the 'My Goals', I think it would be important to link that to your normal check in every day. The information that you provide in your 'My Goals' may be relevant to your check in in terms of exercise, sleep, or how you're feeling today."* [Participant number 7]

- Provide different mode settings to enable tailoring the app 'look and feel' to each user

> *"The only thing I would suggest would be a change of background or some personalized settings so people can really make it their own and keep it engaging."* [Participant number 12]

- Scales used to rate physical activity need to be revised - not considered intuitive

> *"Physical activity levels being levelled from 'great' to 'the worst' doesn't really make a lot of sense. If it were in terms of minutes, that would, I think, be easier to navigate."* [Participant number 10]

Navigation

- Easy to use and intuitive

> *"The navigation around the app was, I felt, really intuitive and really easy."* [Participant number 2]

- More guidance on how to utilize functions not related to check-in

> *"So, I think the main ones obviously that I was using were doing the pain every day, but the other sections I wasn't too familiar with or understanding well. Yes [more guidance on other sections], because I think the reason I stuck to the first one was because I was instructed how to use it."* [Participant number 15]

Perception of overall acceptability

- App was considered a valuable resource for monitoring and managing pain

> *"Yeah, I think the app I would probably use on a daily basis. The app I really, really liked. I don't know whether I'm still able to use it but yeah, if I'm still able to use it now then I'd love to. I think the app is just amazing."* [Participant number 13]

- App readily available using phone

> *"I thought that was so nice and I wish I had that when I was younger… as I said, it would just be so nice because you always have an information source on your phone and you could just use it whenever you needed it."* [Participant number 9]

Metatheme 4: Leveraging Uptake of Digital Tools

Participants provided rich insights into how both these digital tools could be leveraged to extend reach into the community and drive uptake by young people with pain. Key themes included marketing the app to potential users, highlighting key functionalities and features that enable tailoring the app to support their self-management (Textbox 7). Testimonials from current users were perceived as marketing opportunities, especially where the testimonial was linked to a social media account identifying a real-world person who had lived experience of pain. Marketing strategies that included wider promotion through health services or referral by health professionals were perceived by participants as effective ways to increase awareness about these digital tools and their value for self-management of pain. Social media was also advocated as a valuable platform to promote awareness and drive the dissemination of digital tools as well as linking (push) between the app and the (new) website. Some participants suggested promotion via not-for-profit arthritis organizations and educational institutions (schools and universities). To encourage sustained user engagement with digital tools, a few participants recommended a rewards-based design (ie, achieving levels of performance).

Textbox 7. Metatheme 4: leveraging uptake of digital tools.

Market desirable features to the user group

- How the tools can help the user

> *"I think curiosity drives young people. If you make it available and say that it'll help you track your pain and so on, I think that might motivate them for people who want to track or monitor what they did."* [Participant number 14; app]

- Testimonials from users

> *"I think people do often read what other people have to say about it, so I think that definitely [testimonials] would make a difference."* [Participant number 12; app and website]

Multimodal advertising approach

- Referral from a health professional which can be utilized both within and external to the clinical consultation

> *"…having that a doctors' surgeries or having doctors or other allied health, like physios or anything like that, refer it."* [Participant number 8; app and website]

- Social media advertising

> *"I think social media could be a big way. Particularly through Facebook or Instagram, having ads that pop up I think could be a good way."* [Participant number 8]

- Links through other websites

> *"But if it popped up on certain websites, like Beyond Blue or something like that as well, because I feel like a lot of them probably go onto those sorts of spaces to look for support... Yeah, because it makes it seem like it is more reliable and a better source of information if it is popping up on something like the government websites."* [Participant number 1]

- Promotion through educational institutions

> *"I think you've got to put the information out there and I think one way could be through maybe schools…'If you experience this pain, we've got these websites, we've got these apps, there are tools available'. I guess the same could go at a university or education system..."* [Participant number 8]

- Promotion through specific age relevant chronic disease organizations

> *"In WA, I know there's a thing called Camp Freedom, which is an arthritis camp with the JIA, but they talk about quite a lot of different apps, they hand out brochures and that kind of thing. So that is a really helpful way to get in as well, because all these kids are dealing with pain."* [Participant number 13]

Increase longer-term engagement with digital tools by increasing interactivity features

> *"I think there's a lot of really valuable information on the website and in the app, but I think it's that linking to actions and helping to guide people through thought processes, rather than just being information or data tallying. Because it will help engage people for longer with the website and with the services and with the app if they feel like they're making progress via the engagement with those online and app mediums..."* [Participant number 2]

Discussion

Principal Findings

This study aimed to test the acceptability, usability, and need for adaptation of 2 extant digital technologies to support the needs of young Australians with musculoskeletal pain. Overall, the participants' perspectives were positive on the acceptability and usability of both the final prototype pain website and the *iCanCope with Pain* app. In using these digital technologies, participants articulated the critical importance of designs being youth-focused, with colorful, fun interfaces that were relatable to promote young people's engagement in self-management of their musculoskeletal pain. Adaptation of the website prototype was required and informed by rapid cycles of iteration with a focus on improving user engagement. For the app, participant engagement highlighted the value of tailoring capabilities to support individualized pain self-management. Recommendations for app adaptations were modest and primarily related to enhancing functionality (more capacity to personalize the *look and feel*), improved flexibility (check-in time), and capacity to monitor outcomes from goal setting. Participant insights on leveraging these digital technologies to support young people's self-management emphasized the need for wider promotion, health professional *digital prescriptions*, and strategies to ensure longer-term engagement.

Strengths and Limitations

Strengths of this 3-phased mixed methods study included the following: (1) adherence to a user-centered eHealth design roadmap [20] reflecting value specifications, co-design principles, iterative cycles of prototype testing, and identification of factors relevant to implementation; (2) evaluation criteria that predict real-world user engagement [20,36,37]; (3) use of an innovative web-based testing platform *fit for purpose*, allowing remote user testing and reach across geographic barriers; (4) rapid cycles of website prototype design, testing, and iteration; (5) use of an analytics platform to evaluate engagement with capture extended over 1 week (user testing often conducted as a short, single session under laboratory conditions); and (6) mixed methods design, using a larger (user testing) sample, thereby providing rich insights on the acceptability and usability of these digital technologies and informing recommendations on the need for adaptation and suggestions to enable implementation.

Limitations included (1) the potential for gender bias given the dominantly female sample; (2) potential participation bias, with recruitment of young people who felt confident in sharing their experiences or who had directly experienced health services/resource gaps; and (3) although our sample is representative of young people from Australia (a developed country with a high-quality health care system), findings are not necessarily transferable to different health care systems, including those of middle or low-income economies.

Leveraging Digital Technologies to Support Young People's Self-Management of Musculoskeletal Pain

Leveraging these 2 digital tools (app and web-based) for pain that we have previously co-designed, developed, and evaluated [16,30], enabled us to rapidly create a digital test bed in Australia. Advantages included time and cost efficiencies, maximizing the use of current resources, and avoiding unnecessary duplication [19]. Accelerating this phase of our research program to support young Australian people with musculoskeletal pain [1,2,16] is critical [17,41], given the increasing burden imposed by musculoskeletal pain and the lack of an appropriately skilled health workforce [9,15,17]. Digital technologies such as those we have tested in this study provide an innovative approach to improving timely access to credible and practical self-management for young people with musculoskeletal pain and other chronic noncommunicable conditions [2,42,43], regardless of where they live [1]. This is particularly relevant in the current Australian digital landscape given the implementation of the National Digital Health Strategy [44] and the drive for innovative, sustainable digital technologies to support the implementation of health services and self-care of chronic health conditions [29,44].

User-Centered Co-Design Critical to Optimizing Acceptability, Usability, and Engagement

Collectively, the findings from our study highlight the importance of user-centered co-design from inception. Our approach is consistent with evidence derived from our recent systematic review of mobile health technologies for chronic noncommunicable disease management in young people, specifically recommendations on effective implementation [2] and the use of an evidence-based roadmap to increase the uptake and impact of eHealth technologies [20]. Our digital tools were perceived as being youth-focused, with a visual design that elicited positive user emotions and promoted active engagement, outcomes aligned with our primary study aims. The positive perceptions of young people on acceptability, usability, and potential engagement are consistent with qualities of product design that predict real-world user engagement with (app and web-based) eHealth interventions [36]. Furthermore, a recent study on design qualities that may predict user adherence to behavioral eHealth interventions in real-world use, found therapeutic persuasiveness (defined as *the incorporation of persuasive design/behavior change principles*) to be the most robust predictor of adherence (ie, duration of use and number of unique sessions), suggesting the importance of persuasive design and behavior change techniques incorporation during the design and evaluation of digital behavioral interventions [45].

Regarding adaptations, the optimization of the app was primarily related to enhancements in functionality (capacity to personalize the *look and feel*, scales for goal setting and monitoring of goals/outcomes, and time flexibility for check-ins). Use of the app over the week varied with individual participant engagement data, highlighting the value of flexible daily check-ins and tailoring capabilities to support individualized self-reflection and monitoring. A similar approach to testing cultural appropriateness, usability, and the need for adaptation of the *iCanCope with Pain* app has recently been undertaken in Norway [32], with preliminary outcomes indicating high levels of acceptability and usability, the only adaptations being the need for optimizing user interaction of the social support feature [32]. The app is currently being evaluated in a randomized controlled trial with a larger sample with chronic pain and juvenile inflammatory arthritis. Other derivatives of the *iCanCope with Pain* app have also been developed for various conditions in children and adolescents, including self-management of postoperative pain [46], sickle cell pain [47], and juvenile idiopathic arthritis [48], with the latter 2 currently undergoing evaluation [46].

Adaptation of the website prototype was guided by rounds of participant feedback, informing rapid cycles of iteration with a focus on improving user engagement. Recommendations for adaptation are mainly related to greater use of chunked information and call outs (key messages in large font that function as calls to action), optimizing navigation (sticky bars, search functions, and drop-down menus), and improving the information on accessing other services and resources to their support care (what services, what they offer, how to access them). These design recommendations were implemented in the final website prototype in preparation for the current (ongoing) phase of finalizing text and audiovisual content in collaboration with young people. This approach to co-design of content will ensure alignment to Australian user needs and preferences [1,14], while also ensuring consistent messaging and self-management domains between the website and the *iCanCope with Pain* app.

Operationally, the intention for future use of these 2 digital platforms in Australia is to offer complementary digital tools with different functions that can interact bidirectionally (eg, by use of push notifications). Such an approach offers flexibility in supporting both individualized tailored self-management (app), while concurrently providing the capacity for richer audiovisual content and resources specifically codeveloped with and for young Australians (website) and critically, explicitly linked to Australian health services and systems [2]. On the basis of our impact evaluation of the adult pain*HEALTH* website [30], digital platforms can operate as an important health strengthening tool [49], linking consumers/caregivers with their clinicians, services, and systems. We have also demonstrated that this approach promotes more holistic integrated pain care models [17,30] and enables consistency of messaging between consumers and their clinicians (eg, sharing of short targeted audiovisual content during clinical consultations about other people's pain experiences, and how they have implemented positive evidence-based behaviors to improve their pain care). The web-based platform will also include contemporary, evidence-based condition-specific musculoskeletal knowledge (eg, about low back pain and juvenile inflammatory arthritis) with links to best practice pain self-management (eg, making sense of pain, pain education, coping skills and behavioral approaches to pain, encouraging movement, activity and exercise with pain (pacing), appropriate use of medicines) [16,50]. Additional advantages of a web-based platform include the capacity for real-time updating at lower resourcing and cost than app-based technologies, multiple-platform compliance (ie, capacity for use on various devices), easy access, and linking to other entities (eg, consumer, tertiary educational, and health professional bodies). Advantages of the app included empowering young people to take *their health in their hands*, supported by individual tailoring abilities of the app, informed by check-ins, with self-monitoring and self-reflection supporting helpful habits (behavior change) and complementing clinical care.

Conclusions: Insights and the Next Steps

Outcomes from current, ongoing trials of the effectiveness of the *iCanCope with Pain* app will help to inform our decision on the most appropriate approach to evaluate the implementation of these digital technologies in Australian pediatric pain care settings. We envisage the use of a contemporary and flexible evaluation approach that moves beyond traditional effectiveness designs to consider alterative hybrid trial designs [51] or multidimensional and whole-of-system evaluation approaches, such as a *benefits evaluation* [52] that is best aligned to the Australian digital ecosystem and positions these digital tools for real-world implementation [51]. The adaptation of user-centered design and implementation science methods from inception (such as done here) can mitigate the risk of low rates of implementation and the associated research waste [19]. In this context, we gained valuable insights from young people to support and enable implementation, including taking a *whole-of-health* (systems, services, and clinical-level) approach to facilitate real-world dissemination and embedding. These findings are in accordance with recommendations from our recent systematic review of mHealth technologies for noncommunicable chronic disease management in young adults [2].

Digital technologies are positioned to enable the rapid transformation of health care, with their critical role highlighted by the release of the first World Health Organization guideline establishing recommendations on digital interventions for health system strengthening [53], and more recently, local recommendations for transforming health in Australia [54]. The outcomes from this study are, therefore, timely considering the reorientation of health services and system reform toward better integrated management of chronic health conditions [55], healthy aging across the life course [56], and the expectations of young people with chronic musculoskeletal pain for access to evidence-based digital tools to support their self-management [1]. Although we are cautious and thoughtful about the many broader research questions that remain on the evidence for the use of digital technologies in health care innovation (consumer needs and preferences; synergy with current services and systems; capacity to interface with current health systems, services, and workflows; achievement of broader system goals; and stability and sustainability), we suggest that this should not be a reason to accept the current default position in Australia, where many young people with musculoskeletal pain do not have timely access to reliable evidence-based services and resources that promote and support helpful behaviors and improved well-being.

Acknowledgments

This research was supported by grant cofunding awarded from the Government of Western Australia, Department of Health, Health Networks (2018_2436) and Arthritis and Osteoporosis WA, with in-kind support from the School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University. The funding organization had no influence on the study design, collection, analysis, and interpretation of data or on the writing of the manuscript and the decision to submit the manuscript for publication. The authors wish to thank the staff at Improving Outcomes in Child Health through Technology Laboratory, SickKids Hospital, Toronto, Canada, for assistance with providing codes for access to the *iCanCope with Pain* app and for piloting use of the LookBack platform. The authors also acknowledge the support from the Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, and the Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, Ontario, Canada, for the use of APEEE.

Authors' Contributions

HS, JNS, and AMB conceived and led the study. HS, AMB, JNS, and J Cafazzo procured the funding, and HS, JNS, JEJ, AMB, BL, J Chua, CL, QP, and J Cafazzo contributed to the study methods. BL, J Chua, and HS analyzed the web-based phase 1 data, and QP analyzed the phase 2 engagement data (APEEE). JEJ collected and analyzed the phase 3 qualitative interview data and developed the coding framework with review by HS, AMB, JNS, and CL. JNS and CL provided external validation. All authors (HS, JEJ, JNS, J Chua, BL, CL, QP, J Cafazzo, and AMB) contributed to the drafting and review of the final manuscript.

Conflicts of Interest

HS, AMB, J Chua, and BL were involved in the co-design, development, implementation, and evaluation of pain*HEALTH*; JEJ was involved in the evaluation of pain*HEALTH*. JNS, CL, and J Cafazzo were involved in the co-design and evaluation of the *iCanCope with Pain* app; QP, J Cafazzo, JNS, and CL were involved in the co-design, development, and evaluation of APEEE. Grant funding for this project supported the research activities of J Chua, BL, and JEJ and was administered by Curtin University. The remaining authors (HS, JNS, JEJ, J Chua, CL, BL, QP, J Cafazzo, and AMB) have no other conflicts of interest to declare.

Multimedia Appendix 1
Consolidated criteria for reporting qualitative studies: 32-item checklist.
[PDF File (Adobe PDF File), 207 KB - jmir_v22i6e18315_app1.pdf]

Multimedia Appendix 2
Phase 1: website user testing guide.
[PDF File (Adobe PDF File), 159 KB - jmir_v22i6e18315_app2.pdf]

Multimedia Appendix 3
Phase 2: iCanCope with Pain app user testing guide.
[PDF File (Adobe PDF File), 155 KB - jmir_v22i6e18315_app3.pdf]

Multimedia Appendix 4
Phase 3: interview schedule.
[PDF File (Adobe PDF File), 181 KB - jmir_v22i6e18315_app4.pdf]

Multimedia Appendix 5
Phase 1: website user testing criteria-based group-aggregated outcomes.
[PDF File (Adobe PDF File), 192 KB - jmir_v22i6e18315_app5.pdf]

Multimedia Appendix 6

Phase 3: comprehensive summary of metathemes, themes, and subthemes with supporting quotes derived from qualitative interviews.
[PDF File (Adobe PDF File), 534 KB - jmir_v22i6e18315_app6.pdf]

References

1. Slater H, Jordan JE, Chua J, Schütze R, Wark JD, Briggs AM. Young people's experiences of persistent musculoskeletal pain, needs, gaps and perceptions about the role of digital technologies to support their co-care: a qualitative study. BMJ Open 2016 Dec 9;6(12):e014007 [FREE Full text] [doi: 10.1136/bmjopen-2016-014007] [Medline: 27940635]
2. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mhealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
3. Stinson J, White M, Isaac L, Campbell F, Brown S, Ruskin D, et al. Understanding the information and service needs of young adults with chronic pain: perspectives of young adults and their providers. Clin J Pain 2013 Jul;29(7):600-612. [doi: 10.1097/AJP.0b013e31826dce65] [Medline: 23328333]
4. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
5. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
6. Burns J, Birrell E. Enhancing early engagement with mental health services by young people. Psychol Res Behav Manag 2014;7:303-312 [FREE Full text] [doi: 10.2147/PRBM.S49151] [Medline: 25473320]
7. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
8. Analysis & Policy Observatory (APO). 2011. The High Price of Pain: The Economic Impact of Persistent Pain in Australia URL: https://apo.org.au/node/3054 [accessed 2020-02-16]
9. Musculoskeletal Australia (MSK). 2013. A Problem Worth Solving: The Rising Cost of Musculoskeletal Conditions in Australia URL: https://www.msk.org.au/wp-content/uploads/2018/07/APWS-PLS.pdf [accessed 2020-02-16]
10. O'Sullivan PB, Straker LM, Smith A, Perry M, Kendall G. Carer experience of back pain is associated with adolescent back pain experience even when controlling for other carer and family factors. Clin J Pain 2008;24(3):226-231. [doi: 10.1097/AJP.0b013e3181602131] [Medline: 18287828]
11. King S, Chambers CT, Huguet A, MacNevin RC, McGrath PJ, Parker L, et al. The epidemiology of chronic pain in children and adolescents revisited: a systematic review. Pain 2011 Dec;152(12):2729-2738. [doi: 10.1016/j.pain.2011.07.016] [Medline: 22078064]
12. Gobina I, Villberg J, Välimaa R, Tynjälä J, Whitehead R, Cosma A, et al. Prevalence of self-reported chronic pain among adolescents: evidence from 42 countries and regions. Eur J Pain 2019 Feb;23(2):316-326. [doi: 10.1002/ejp.1306] [Medline: 30098106]
13. Swain MS, Henschke N, Kamper SJ, Gobina I, Ottová-Jordan V, Maher CG. An international survey of pain in adolescents. BMC Public Health 2014 May 13;14:447 [FREE Full text] [doi: 10.1186/1471-2458-14-447] [Medline: 24885027]
14. Slater H, Jordan J, Chua J, Schutze R, Briggs A. Arthritis WA. 2016. Painful Transitions: Young People's Experiences of Living With Persistent Pain, Their Interactions With Health Services and Their Needs and Preferences for Pain Management Including Digital Technologies URL: http://www.arthritiswa.org.au/wp-content/uploads/2016/10/BeRBvB_FINAL-Painful-Transitions-report.pdf [accessed 2020-02-16]
15. PainAustralia. 2010. National Pain Strategy URL: https://www.painaustralia.org.au/improving-policy/national-pain-strategy [accessed 2020-02-16]
16. Stinson JN, Lalloo C, Harris L, Isaac L, Campbell F, Brown S, et al. iCanCope with pain: user-centred design of a web- and mobile-based self-management program for youth with chronic pain based on identified health care needs. Pain Res Manag 2014;19(5):257-265 [FREE Full text] [doi: 10.1155/2014/935278] [Medline: 25000507]
17. Slater H, Dear BF, Merolli MA, Li LC, Briggs AM. Use of ehealth technologies to enable the implementation of musculoskeletal models of care: evidence and practice. Best Pract Res Clin Rheumatol 2016 Jun;30(3):483-502 [FREE Full text] [doi: 10.1016/j.berh.2016.08.006] [Medline: 27886943]
18. Hogg MN, Gibson S, Helou A, DeGabriele J, Farrell MJ. Waiting in pain: a systematic investigation into the provision of persistent pain services in Australia. Med J Aust 2012 Apr 2;196(6):386-390. [doi: 10.5694/mja12.10140] [Medline: 22471539]
19. Higgins KS, Tutelman PR, Chambers CT, Witteman HO, Barwick M, Corkum P, et al. Availability of researcher-led ehealth tools for pain assessment and management: barriers, facilitators, costs, and design. Pain Rep 2018 Sep;3(Suppl 1):e686 [FREE Full text] [doi: 10.1097/PR9.0000000000000686] [Medline: 30324177]

20. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of ehealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
21. Briggs AM, Slater H, Bunzli S, Jordan JE, Davies SJ, Smith AJ, et al. Consumers' experiences of back pain in rural western Australia: access to information and services, and self-management behaviours. BMC Health Serv Res 2012 Oct 11;12:357 [FREE Full text] [doi: 10.1186/1472-6963-12-357] [Medline: 23057669]
22. Briggs AM, Slater H, Smith AJ, Parkin-Smith GF, Watkins K, Chua J. Low back pain-related beliefs and likely practice behaviours among final-year cross-discipline health students. Eur J Pain 2013 May;17(5):766-775. [doi: 10.1002/j.1532-2149.2012.00246.x] [Medline: 23139051]
23. Bunzli S, Watkins R, Smith A, Schütze R, O'Sullivan P. Lives on hold: a qualitative synthesis exploring the experience of chronic low-back pain. Clin J Pain 2013 Oct;29(10):907-916. [doi: 10.1097/AJP.0b013e31827a6dd8] [Medline: 23370072]
24. Slater H, Briggs AM, Bunzli S, Davies SJ, Smith AJ, Quintner JL. Engaging consumers living in remote areas of western Australia in the self-management of back pain: a prospective cohort study. BMC Musculoskelet Disord 2012 May 11;13:69 [FREE Full text] [doi: 10.1186/1471-2474-13-69] [Medline: 22578207]
25. Slater H, Davies SJ, Parsons R, Quintner JL, Schug SA. A policy-into-practice intervention to increase the uptake of evidence-based management of low back pain in primary care: a prospective cohort study. PLoS One 2012;7(5):e38037 [FREE Full text] [doi: 10.1371/journal.pone.0038037] [Medline: 22662264]
26. Slater H, Briggs AM, Smith AJ, Bunzli S, Davies SJ, Quintner JL. Implementing evidence-informed policy into practice for health care professionals managing people with low back pain in Australian rural settings: a preliminary prospective single-cohort study. Pain Med 2014 Oct;15(10):1657-1668. [doi: 10.1111/pme.12351] [Medline: 24433536]
27. Davies SJ, Hayes C, Quintner JL. System plasticity and integrated care: informed consumers guide clinical reorientation and system reorganization. Pain Med 2011 Jan;12(1):4-8. [doi: 10.1111/j.1526-4637.2010.01016.x] [Medline: 21143757]
28. Briggs AM, Bragge P, Slater H, Chan M, Towler SC. Applying a health network approach to translate evidence-informed policy into practice: a review and case study on musculoskeletal health. BMC Health Serv Res 2012 Nov 14;12:394 [FREE Full text] [doi: 10.1186/1472-6963-12-394] [Medline: 23151082]
29. The COAG Health Council. 2017. National Strategic Framework for Chronic Conditions URL: https://www.coaghealthcouncil.gov.au/Portals/0/National%20Strategic%20Framework%20for%20Chronic%20Conditions.pdf [accessed 2020-01-26]
30. Slater H, Jordan JE, Houlding E, Davies S, Milne GJ, Burley M, et al. WA Health, Government of Western Australia. 2019. painHEALTH Evaluation Report Executive Summary URL: https://ww2.health.wa.gov.au/Articles/J_M/Musculoskeletal-Health-Network [accessed 2019-11-30]
31. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar 8;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]
32. Grasaas E, Fegran L, Helseth S, Stinson J, Martinez S, Lalloo C, et al. iCanCope with pain: cultural adaptation and usability testing of a self-management app for adolescents with persistent pain in Norway. JMIR Res Protoc 2019 Jun 3;8(6):e12940 [FREE Full text] [doi: 10.2196/12940] [Medline: 31162132]
33. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
34. Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Örebro musculoskeletal pain screening questionnaire. Spine (Phila Pa 1976) 2011 Oct 15;36(22):1891-1895. [doi: 10.1097/BRS.0b013e3181f8f775] [Medline: 21192286]
35. Beales DJ, Smith AJ, O'Sullivan PB, Straker LM. Low back pain and comorbidity clusters at 17 years of age: a cross-sectional examination of health-related quality of life and specific low back pain impacts. J Adolesc Health 2012 May;50(5):509-516. [doi: 10.1016/j.jadohealth.2011.09.017] [Medline: 22525116]
36. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided ehealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
37. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based ehealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
38. Pham Q, Graham G, Lalloo C, Morita PP, Seto E, Stinson JN, et al. An analytics platform to evaluate effective engagement with pediatric mobile health apps: design, development, and formative evaluation. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11447 [FREE Full text] [doi: 10.2196/11447] [Medline: 30578179]
39. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded. Thousand Oaks, CA: Sage Publications; 1998.
40. Patton MC. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: Sage Publications; 2014.

41. Briggs AM, Chan M, Slater H. Models of care for musculoskeletal health: moving towards meaningful implementation and evaluation across conditions and care settings. Best Pract Res Clin Rheumatol 2016 Jun;30(3):359-374. [doi: 10.1016/j.berh.2016.09.009] [Medline: 27886937]
42. World Health Organization. Atlas of eHealth Country Profiles: the Use of eHealth in Support of Universal Health Coverage: Based on the Findings of the Third Global Survey on eHealth 2015. Geneva, Switzerland: World Health Organization; 2015.
43. Deloitte Access Economic. 2019. The Cost of Pain in Australia URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/Economics/deloitte-au-economics-cost-pain-australia-040419.pdf [accessed 2020-01-06]
44. Australia's National Digital Health Agency. 2017. Australia's National Digital Health Strategy - Safe, Seamless and Secure: Evolving Health and Care to Meet the Needs of Modern Australia URL: https://conversation.digitalhealth.gov.au/australias-national-digital-health-strategy [accessed 2019-12-01]
45. Baumel A, Yom-Tov E. Predicting user adherence to behavioral ehealth interventions in the real world: examining which aspects of intervention design matter most. Transl Behav Med 2018 Sep 8;8(5):793-798. [doi: 10.1093/tbm/ibx037] [Medline: 29471424]
46. Birnie KA, Campbell F, Nguyen C, Lalloo C, Tsimicalis A, Matava C, et al. iCanCope PostOp: user-centered design of a smartphone-based app for self-management of postoperative pain in children and adolescents. JMIR Form Res 2019 Apr 22;3(2):e12028 [FREE Full text] [doi: 10.2196/12028] [Medline: 31008704]
47. Palermo TM, Zempsky WT, Dampier CD, Lalloo C, Hundert AS, Murphy LK, et al. iCanCope with sickle cell pain: design of a randomized controlled trial of a smartphone and web-based pain self-management program for youth with sickle cell disease. Contemp Clin Trials 2018 Nov;74:88-96 [FREE Full text] [doi: 10.1016/j.cct.2018.10.006] [Medline: 30316999]
48. Lalloo C, Harris L, Hundert A, Berard R, Cafazzo J, Connelly M. Pilot randomized controlled trial of the iCanCope pain self-management application for adolescents with juvenile idiopathic arthritis. Rheumatology 2020 In press.
49. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
50. Stinson J, Connelly M, Kamper SJ, Herlin T, Toupin April K. Models of care for addressing chronic musculoskeletal pain and health in children and adolescents. Best Pract Res Clin Rheumatol 2016 Jun;30(3):468-482. [doi: 10.1016/j.berh.2016.08.005] [Medline: 27886942]
51. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mhealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
52. Biggs JS, Willcocks A, Burger M, Makeham MA. Digital health benefits evaluation frameworks: building the evidence to support Australia's national digital health strategy. Med J Aust 2019 Apr;210(Suppl 6):S9-11. [doi: 10.5694/mja2.50034] [Medline: 30927475]
53. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019.
54. CSIRO. 2018. Future of Health, Shifting Australia's Focus From Illness Treatment to Health and Wellbeing Management URL: https://www.csiro.au/en/Showcase/futureofhealth [accessed 2019-09-03]
55. Briggs AM, Persaud JG, Deverell ML, Bunzli S, Tampin B, Sumi Y, et al. Integrated prevention and management of non-communicable diseases, including musculoskeletal health: a systematic policy analysis among OECD countries. BMJ Glob Health 2019;4(5):e001806 [FREE Full text] [doi: 10.1136/bmjgh-2019-001806] [Medline: 31565419]
56. World Health Organization. World Report on Ageing and Health. Geneva, Switzerland: World Health Organization; 2015.

Abbreviations

APEEE: analytics platform to evaluate effective engagement
eHealth: electronic health
ÖMPSQ-SF: Örebro Musculoskeletal Pain Screening Questionnaire—Short Form

---

Edited by G Eysenbach; submitted 20.02.20; peer-reviewed by T Muto, A Tsimicalis, C Varsi; comments to author 20.03.20; revised version received 06.04.20; accepted 10.04.20; published 05.06.20.

Please cite as:
Slater H, Stinson JN, Jordan JE, Chua J, Low B, Lalloo C, Pham Q, Cafazzo JA, Briggs AM
Evaluation of Digital Technologies Tailored to Support Young People's Self-Management of Musculoskeletal Pain: Mixed Methods Study
J Med Internet Res 2020;22(6):e18315
URL: https://www.jmir.org/2020/6/e18315
doi:10.2196/18315
PMID:32442143

Original Paper

Coach-Facilitated Web-Based Therapy Compared With Information About Web-Based Resources in Patients Referred to Secondary Mental Health Care for Depression: Randomized Controlled Trial

Sarah MacLean[1], BA, BSocSci, MA; Daniel J Corsi[2,3,4,5], PhD; Sadie Litchfield[6], MSW; Julia Kucharski[7], MD; Kira Genise[8], BSc; Zeynep Selaman[9], MD; Valerie Testa[2,10], BA, BEd; Simon Hatcher[2,9], MBBS, MD

[1]School of Journalism and Communication, Carleton University, Ottawa, ON, Canada
[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[3]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
[4]Better Outcomes Registry & Network Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada
[5]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada
[6]Prevention and Rehabilitation, University of Ottawa Heart Institute, Ottawa, ON, Canada
[7]Department of Psychiatry, The Ottawa Hospital, Ottawa, ON, Canada
[8]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[9]Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada
[10]Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Simon Hatcher, MBBS, MD
Clinical Epidemiology Program
Ottawa Hospital Research Institute
1919 Riverside Drive
Suite 406
Ottawa, ON, K1H 7W9
Canada
Phone: 1 613 737 8899
Email: shatcher@toh.ca

Abstract

Background: Depression is a common mental disorder with a high social burden and significant impact on suicidality and quality of life. Treatment is often limited to drug therapies because of long waiting times to see psychological therapists face to face, despite several guidelines recommending that psychological treatments should be first-line interventions for mild to moderate depression.

Objective: We aimed to evaluate, among patients on a waitlist to receive secondary mental health care services for depression, how effective coach-guided web-based therapy (*The Journal*) is, compared with an information-only waitlist control group, in reducing depression symptoms after 12 weeks.

Methods: We conducted a randomized controlled trial with 2 parallel arms and a process evaluation, which included interviews with study participants. Participants assigned to the intervention group received 12 weeks of web-based therapy guided by a coach who had a background in social work. Patients in the control group receive a leaflet of mental health resources they could access. The primary outcome measure was a change in depression scores, as measured by the Patient-Health Questionnaire (PHQ-9).

Results: A total of 95 participants were enrolled (intervention, n=47; control, n=48). The mean change in PHQ-9 scores from baseline to week 12 was −3.6 (SD 6.6) in the intervention group and −3.1 (SD 6.2) in the control group, which was not a statistically significant difference with a two-sided alpha of .05 ($t_{91}=-0.37$; *P*=.72, 95% CI −3.1 to 2.2). At 12 weeks, participants in the intervention group reported higher health-related quality of life (mean EuroQol 5 dimensions visual analogue scale [EQ-5D-VAS] score 66.8, SD 18.0) compared with the control group (mean EQ-5D VAS score 55.9, SD 19.2; $t_{84}=-2.73$; *P*=.01). There were no statistically significant differences between the two groups in health service use following their initial consultation with a psychiatrist. The process evaluation showed that participants in the intervention group completed a mean of 5.0 (SD 2.3) lessons in *The Journal* and 8.8 (SD 3.1) sessions with the coach. Most participants (29/47, 62%) in the intervention group who completed the full dose of the intervention, by finishing 6 or more lessons in *The Journal*, were more likely to have a clinically important reduction in depressive symptoms at 12 weeks compared with the control group (X^2_1=6.3; *P*=.01, Φ=0.37). Participants who completed the interviews reported that the role played by the coach was a major factor in adherence to the study intervention.

Conclusions: The results demonstrate that the use of guided web-based therapy for the treatment of depression is not more effective than information-only waitlist control. However, it showed that the coach has the potential to increase adherence and engagement with web-based depression treatment protocols. Further research is needed on what makes the coach effective.

Trial Registration: ClinicalTrials.gov: NCT02423733; https://clinicaltrials.gov/ct2/show/NCT02423733

(J Med Internet Res 2020;22(6):e15001) doi:10.2196/15001

KEYWORDS

major depressive disorder; secondary care; randomized controlled trial; telemedicine; digital health technologies; Canada

Introduction

Background

Depression is a common mental disorder [1] with a high social burden [2] and significant impact on suicidality [3] and quality of life. Most treatment for depression occurs in primary care. In secondary care, treatment is often limited to drug therapies, in part, due to long waiting lists to see psychological therapists face-to-face. This is contrary to recommendations about the importance of nondrug therapies by the National Institute for Health and Care Excellence (NICE), based in the United Kingdom, and other institutions [4]. In secondary and tertiary mental health care centers in Ontario, at the time of this trial, the waiting time to be seen by a psychiatrist for depression was between 9 months and 1 year.

Use of Electronic Therapies in Treating Depression

There is evidence that web-based therapies can reduce the symptoms of depression [5]. Randomized controlled trials (RCTs) have demonstrated the effectiveness of web-based cognitive behavior therapy (CBT) [6,7], problem-solving therapy [8], interpersonal therapy [9], and psychodynamic therapy. Computerized CBT is also recommended in the NICE guidelines for the treatment of mild to moderate depression [4]. However, most of the RCTs of web-based therapies have been conducted in community samples, often recruited from the internet. These populations are self-selecting and, although their scores on depression rating scales may be comparable with clinical populations, they often differ in terms of comorbidity, duration of symptoms, and impact on daily functioning.

Guided Electronic Therapy in the Treatment of Depression

The delivery of web-based therapy as a treatment for depression can be performed in one of two ways: through the use of supports to assist patients through the web-based therapy (guided model) or through the self-help use of computerized treatment (unguided model). In the guided model, patients are provided support as they progress through web-based therapy. In some cases, highly trained clinicians have been used to fill this role [10]; however, their advanced training is costly and does not appear to provide any added benefit. For instance, despite the use of a clinician in the delivery of web-based therapy for the treatment of depression and anxiety among young adults, Dear et al [11] did not find any significant differences in symptomatology when compared with the unguided condition. One solution that has been implemented to mitigate this issue is the use of coaches to support patients progressing through web-based therapy, answering both technical questions about how the program works as well as providing support and encouragement. To date, coaches have included both students [12] and licensed professionals from a variety of backgrounds, including psychology [13], counseling [14], and social work [15,16]. The components of effective coaching are still uncertain [17], with most of the literature emphasizing the technical aspects of an internet intervention. It is not clear whether the professional background of the coach, the frequency of contact, or the content of the coaching sessions affects treatment outcomes.

Several systematic reviews have demonstrated that internet-based interventions for depression have effect sizes that are comparable with face-to-face interventions, whereas unguided interventions have smaller effect sizes [7,18,19]. However, the evidence is inconsistent, with head-to-head comparisons of guided versus unguided interventions showing mixed results. The largest study of guided web-based therapy compared with usual care has been the Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) study [20], which was a UK-based RCT of 691 patients with depression in primary care. Participants were randomized to one of three treatment groups: (1) usual care; (2) *Beating the Blues*, a guided commercial web-based program; or (3) *MoodGYM*, a guided free web-based program. The study found no difference in depression outcomes between the three groups. Similarly, Kenter et al [21] compared student-assisted internet-based problem-solving therapy with an information-only waitlist control and found minimal differences in depression severity (Cohen *d*=0.07), as measured by the Centre for Epidemiology Studies Depression scale.

Adherence to Web-Based Therapy Treatment Protocols

Web-based interventions for depression struggle with engagement and adherence to treatment protocols. For example, in the *Beating the Blues* and *MoodGYM* arms of the REEACT trial, participants completed only a median of 1 to 2 sessions and received only 6 min of technical support time, 5 emails, and almost no text messages from the telephone support workers. Further, approximately 1 in 5 participants randomized to either of the web-based therapy conditions did not access the programs at all. Similar problems were encountered in the study conducted by Kenter et al [21], with only 36% of participants receiving an adequate dosage of the intervention, defined by the authors as 4 of 5 lessons of web-based therapy. From a public health perspective, low adherence to web-based therapies dilutes their effectiveness. However, it is unclear if, for individual patients, adhering to a complete course of web-based therapy is better than control treatments.

This Study

We have previously reported on an RCT of *The Journal* comparing guided web-based therapy with an information-only waitlist control in patients with depression referred to secondary mental health services in New Zealand [22]. Participants were recruited face-to-face during their triage visit at community mental health centers, and the study found no difference in depression or service use after 12 weeks. However, all participants in this study were also receiving mental health care from clinicians, which may have outweighed any effect of the web-based therapy. In this study, we report on an RCT of *The Journal* [23], facilitated by a coach, compared with an information-only waitlist control group for the treatment of depression in patients referred to secondary mental health services in Canada. In this setting, access is a significant issue, with patients often waiting for over a year to access mental health services. During this time, patients on the waiting list do not receive any other mental health care other than routine follow-up from their family physician; therefore, the provision of web-based therapy could potentially be used as an alternative to referral to specialized mental health services. This study gave us an opportunity to refine the study intervention. In contrast to the New Zealand study, we opted to use an information-only control group as it is a low-cost, low-risk alternative that could easily be implemented in clinical practice. Similarly, we were able to refine the coaching aspect of the intervention. We hypothesized that, after 12 weeks of treatment, participants receiving coach-guided web-based therapy would experience a greater reduction in depressive symptoms and health service use.

Methods

Trial Design

The design of the trial was an RCT with two parallel groups. This trial has been reported according to the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-EHEALTH; Multimedia Appendix 1).

Recruitment

Potential participants were patients referred to the Royal Ottawa Mental Health Centre (ROMHC, Ottawa, Canada) with symptoms of depression or dysthymia who were on a waiting list for treatment in the following psychiatric programs: Mood and Anxiety, Geriatric Psychiatry and Youth Psychiatry. The ROMHC has 284 inpatient beds and acts as a specialized mental health facility for residents of communities across Eastern Ontario. These programs are aimed at treating people with complex and serious mental illnesses that are often resistant to treatment. Patients are referred to treatment in these programs directly by their family physicians. The first appointment in the programs is with a psychiatrist who then decides on a treatment plan with the patient. At the time of this clinical trial, the waitlist for the Mood and Anxiety Program was between 9 months and 1 year.

Eligibility criteria for participation in the trial are outlined in Textbox 1. Potentially eligible patients were contacted using the following methods: patients who had completed a consent to be contacted for research as part of their referral documentation were contacted by telephone. Those who did not complete the consent to be contacted for the research section of the referral were contacted by mail, as per institutional policies. Interested patients were preliminarily screened by a research assistant for eligibility to participate in the study. Eligible patients were then asked to attend a face-to-face consenting appointment with a research assistant. A minority of patients who could not travel to the clinic were asked to consent via mail (n=3).

Textbox 1. Participant eligibility criteria.

Inclusion criteria:

- 16 years of age or older.
- Referred and triaged to the Mood and Anxiety, Youth Psychiatry, or Geriatric Psychiatry Program at the Royal Ottawa Mental Health Centre for any depressive symptoms.
- Willing to attend electronic therapy sessions for up to 12 weeks.
- Able and willing to provide informed consent.
- Willing to be randomized.
- Willing to comply with all study procedures.

Exclusion criteria:

- Is unable to read or write in English.
- Does not have an Ontario Health Insurance Plan number.
- Has cognitive impairments that render them unable to use a computer.
- There is another participant enrolled in the study who lives at their address.

Interventions

The Journal

The Journal [23] is an evidence-based free web-based therapy program developed in New Zealand for the self-management of depression that utilizes the cognitive behavioral techniques of behavioral activation and problem solving (Multimedia Appendix 2). The problem-solving approach was derived from a large RCT of face-to-face problem-solving used in people who presented to emergency departments with intentional self-harm [24].

As described in Table 1, there are a total of 9 modules in *The Journal*, 6 of which must be done for patients to complete the program. Participants progressed through the web-based therapy as follows: (1) positivity module (1 lesson); (2) lifestyle modules (4 lessons, 1 of which must be completed); and (3) problem-solving module (4 lessons). At the start of each lesson, users are asked to watch a video featuring Sir John Kirwan, a former All-Blacks rugby player, who has been very public about his own struggles with depression in an effort to reduce stigmatization.

Table 1. Breakdown of participant progress through The Journal.

Lessons	Description	Tasks to be completed by participants
Positivity module	Learn the importance of staying positive and planning regular activities that they enjoy	• Watch video on staying positive; • Select 2 enjoyable activities; • Select dates to complete activities.
Lifestyle module[a]		
Eating right	Explore the link between diet and mood.	• Watch video on eating right; • Browse and select a healthy recipe; • Create a shopping plan.
Getting active	Review benefits of being active on mood.	• Watch video on getting active; • Pick 2 activities to complete; • Make a plan for getting active.
Learning to relax	Highlights the importance of stress management.	• Watch video on learning to relax; • Practice relaxation and breathing; • Make a plan for relaxing exercises.
Sleeping better	Discuss the importance of good sleep habits to mood.	• Watch video on sleeping better; • Set a nighttime routine; • Set a morning routine; • Keep a sleep diary; • Make a plan to practice sleep hygiene.
Problem-solving module		
Identify problems	Learn how depression impacts problem-solving abilities.	• Watch video on identifying problems; • Create a problem list; • Pick a problem to work on; • Define the problem; • Make a plan to create a problem list and statement.
Find solutions	Explore how to use both logical and creative parts of the brain to brainstorm problem solutions.	• Watch video on brainstorming solutions; • Create a solutions list; • Select a solution to implement; • Evaluate solutions; • Make a plan to list and evaluate solutions.
Create a plan	Review how to create SMART[b] plan.	• Watch video on brainstorming solutions; • Review the selected solution to make sure it is SMART; • Write a detailed step-by-step plan; • Review plan.
Review your plan	Highlights the importance of assessing progress and updating the SMART plan.	• Watch video on reviewing the SMART plan; • Review progress on plan; • Revise plan as needed; • Complete self-test.

[a] Participants are only required to complete 1 of the 4 lifestyle lessons.

[b] SMART: specific, measurable, achievable, relevant, time-bound.

Study Intervention

Both groups received usual care while on the waitlist, which included management by a family physician and use of community resources, such as access to distress center lines and counseling services. Once participants are called off the waitlist, they receive an initial appointment with a psychiatrist, at which point adjustments are made to the patient's care plan. They are then referred back to their family physician, or they receive further psychiatric treatment (eg, regular care from a psychiatrist, social worker, nurse, occupational therapist, and so on).

Following the consent appointment, participants were randomized to one of two treatment groups. In addition to usual care, participants assigned to the control group received an information leaflet with web-based resources, including *The Journal*, and told that they could decide for themselves the best way to use this information while on the waitlist. Participants had been previously informed by their clinical team that the estimated wait time to be seen by a psychiatrist was between 9 and 12 months.

For those participants assigned to the intervention group, the intervention consisted of the following:

1. An information leaflet of web-based depression resources.
2. An invitation to use *The Journal*.
3. 12-weekly telephone coaching sessions with a coach (SL), who had a guideline script for each coaching session, reinforced the topic of each lesson in *The Journal*, helped identify and support participants in goal setting and the techniques of problem-solving. Each session lasted between 30 and 60 min. The coach had a background in social work and received weekly supervision from the principal investigator (SH).
4. Text message or email contact between appointments, as per the participant's preference.

Outcomes

Primary Outcome

The primary outcome measure was the Patient Health Questionnaire (PHQ-9) [25], a 9-item questionnaire that assesses the severity of depression symptoms experienced within the preceding 2 weeks. Participants are asked to rate each symptom of depression on a Likert scale from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 (minimal depression) to 27 (severe depression). The PHQ-9 has strong methodological properties, with an internal consistency of 0.89 and strong test-retest reliability [26]. Increasing scores on the PHQ-9 have also been found to be correlated with deteriorating scores on all 6 subscales of the Medical Outcomes Survey Short Form-20 [27]. The PHQ-9 was selected not only for its strong psychometric properties but also for its commonality. The PHQ-9 is often used as a screening tool for major depressive disorder in primary care practice [28].

Secondary Outcomes

Suicidal thoughts were assessed by question 9 of the PHQ-9, in which respondents were asked *Over the last 2 weeks how often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way*? [26]. This variable was dichotomized as follows: participants who responded *Not at All* (0) were categorized as *no* (0), and participants who reported any degree of suicidality (1, 2, or 3) were categorized as *yes* (1).

Health-related quality of life was assessed using the EuroQol-5 dimension (3 levels) questionnaire (EQ-5D-3L). This is a 5-item questionnaire that assesses health-related quality of life, including mobility, self-care, ability to participate in one's usual activities, pain or discomfort, and anxiety or depression. The EQ-5D-3L asks participants to assess their health-related quality of life on a 3-point scale from no dysfunction to extreme dysfunction, with the following response categories:

- Level 1: indicating no problem.
- Level 2: indicating some problems.
- Level 3: indicating extreme problems.

The EQ-5D-3L is then able to define a unique health state based on the responses to each of the 5 dimensions of health described above. Respondents fall into 1 of 243 different health states, depending on their responses to the questionnaire [28]. For instance, an overall score of 11111 indicates no problems in any of the 5 health dimensions, whereas a score of 12312 indicates that a respondent has no problems with mobility, some problems with washing or dressing, extreme problems with doing usual activities, no pain or discomfort, and some anxiety or depression. The measure also includes a visual analogue scale (VAS), which asks participants to evaluate their overall health on a scale from 0 to 100. The EQ-5D-3L has strong psychometric properties and has been found to be moderately to highly correlated with other measures of impairment and disability [29,30].

Service use was measured using data extracted from participants' electronic medical records (EMR), including, time to first consultation appointment at the ROMHC and the total number of outpatient mental health follow-up appointments after the first consultation completed at the ROMHC. These measures were administered as shown in Table 2.

Baseline assessments were administered in-person following the consent appointment, and all other time point assessments were conducted by telephone either by the coach (intervention group) or a research assistant (control group). Patients who missed appointments or were lost to follow-up were also sent questionnaires by mail. Service use at the ROMHC was obtained from the EMR of each participant. Five participants in the control group were excluded ad hoc to prevent confounding as they accessed *The Journal* during the treatment period.

Table 2. Outcome measures and timing of assessments.

Variable	Outcome measure	Time point
Primary outcome		
Depressive symptoms	PHQ-9[a]	Baseline, week 2, week 6, week 12
Secondary outcomes		
Suicidal thoughts	PHQ-9 Q9[b]	Baseline, week 2, week 6, week 12
Health-related quality of life	EQ-5D-3L[c], EQ-5D-VAS[d]	Baseline, week 6, week 12
Health service use	ROMHC EMR[e]	One year following the initial consultation appointment at the ROMHC

[a] PHQ-9: Patient Health Questionnaire.

[b] PHQ-9 Q9: Patient Health Questionnaire question 9.

[c] EQ-5D-3L: EuroQol 5 dimensions (3 levels) questionnaire.

[d] EQ-5D-VAS: EuroQol 5 dimensions visual analogue scale.

[e] ROMHC EMR: Royal Ottawa Mental Health Center electronic medical record.

Sample Size

Based on the previous studies that used the PHQ-9 as their primary outcome measure, we expected the mean pretreatment score to be 17.0 (SD 4.0). To detect a difference in PHQ-9 scores between the two groups of at least 3 points, an established minimal clinically important difference [31], we would need 44 participants in each group with a 2-sided alpha of .05, and a power of 80.0% (effect size of 0.6). Allowing for a 25.0% dropout rate, we aimed to recruit a total of 110 participants.

Randomization

Randomization was completed by the Ottawa Methods Centre at the Ottawa Hospital Research Institute, with allocations kept in sequential sealed envelopes at the study base. Participants were randomized in a 1:1 allocation, and there were no restrictions. After providing consent, participants were randomized by a research assistant according to the allocation in the sealed envelopes.

Blinding

Owing to the nature of the intervention, neither participants nor study staff were blinded to the treatment allocation. Outcome assessments were collected by delegated study staff who were not blinded to the treatment allocation.

Statistical Analyses

Group differences in demographic and pretreatment measures were analyzed using independent samples *t* tests. Changes in participants' scores from pretreatment to follow-up at 12 weeks were assessed using repeated measures analysis of variance with mixed linear modeling to account for missing variables. This model included the following variables: treatment group (control or intervention), PHQ-9 scores at 4 different time points (baseline, week 2, week 6, and week 12), gender (male or female), and age. PHQ-9 scores were entered as within-subject variables, treatment group as between-subject factors, and age and gender as covariates. Statistical analyses were conducted using the general linear modeling repeated measures procedure in IBM SPSS Statistics 25 for Windows. To assess significant differences in outcome measures, the last observations for the PHQ-9, EQ-5D-3L and EQ-5D-VAS were carried forward for the 6- and 12-week time points.

Differences in suicidal thoughts between the two groups were assessed using an independent samples *t* test. Differences in proportions of service use were assessed using chi-square tests. In addition, we also assessed whether participants experienced a clinically important reduction in depression symptoms defined as a PHQ-9 score of 9 or less or a 50.0% reduction in scores [31]. This was described using percentages and frequencies, and chi-square was used to assess the significance of the differences in proportions. Assessments of normality were completed using the Shapiro-Wilk test for continuous data, and the Mann-Whitney U test for nonparametric data.

Process Evaluation

As per the recommendations outlined in the Medical Research Council's guidelines for the assessment of complex interventions [32], we conducted a process evaluation to assess the context, implementation, and mechanisms of impact of the study intervention. The process evaluation outcome measures are outlined in Table 3. We also conducted semistructured interviews with participants within 6 months of study completion to assess their experience of the study intervention, including access to and functionality of *The Journal*, the therapeutic content and value of *The Journal*, and the experience of working with a coach. Interviewees were identified using a purposive sampling approach and were stratified to include participants in both arms of the trial and varying levels of engagement (eg, not engaged, moderately engaged, and highly engaged). Interviews were conducted until data saturation was reached (n=15).

Table 3. Process evaluation outcome measures.

Evaluation critierion and outcome measure	Description
Context	
Facilitators of and barriers to study completion	- Qualitative interviews with participants
Implementation	
Reach	- Total number of participants reached - Comparison of sample to Ontarian and Canadian populations.
Fidelity	- Mean length of weekly coaching calls - Mean number of contacts with the coach - Mean number of weekly coaching sessions completed - Mean number of lessons in *The Journal*[a]
Dose	- Total number of participants to complete 6 lessons in *The Journal*
Mechanisms of impact	
Role of the coach	- Qualitative interviews with participants

[a] All usage data from *The Journal* were assessed via participant self-report.

Individual interviews were conducted by two female University of Ottawa Department of Psychiatry residents (JK and ZS) who were independent of the research team that conducted the RCT. Following transcription, two independent coders (JK and KG) analyzed the material using a thematic, grounded theory approach. The coding took place in two stages, with coders meeting during the first stage to discuss emergent codes, reconcile definitions, and compare coding rationales. The two coders had not been involved in delivering the treatment or any of the previous study contacts.

Ethics

The study received approval from the Royal Research Ethics Board (Protocol 2014001). All participants provided informed consent before participation in both the RCT and qualitative interviews.

Results

Participants

Recruitment for this study took place over 11 months, from May 2015 to April 2016, and a total of 1316 patients were preliminarily screened for eligibility by examining their referral documentation (Figure 1). Of these, 45.9% (605/1316) could not be reached to complete a full screening for eligibility; 40.4% (532/1316) could be contacted were not eligible to participate, and 13.6% (179/1316) were eligible and could be contacted.

Of the 45.9% (605/1316) of patients who could not be contacted to complete a full eligibility screen, 25.4% (154/605) provided consent to be approached about research studies but did not reply to telephone or mail invitations; 63.1% (382/605) were invited to discuss the study by mail with no response, and 11.3% (69/606) had incomplete mailing information listed on their referral documents.

Of the 40.4% (532/1316) of patients who could be contacted but were ineligible to participate, 84.9% (452/532) were referred for a reason other than depressive symptoms; 5.6% (30/532) had cognitive impairments rendering them unable to use a computer; 4.5% (24/532) were no longer on the waitlist at the time of screening; 2.6% (14/532) were under 16 years of age at the time of screening; 1.12% (6/532) were unable to read and write in English; and, 1.12% (6/532) did not have a valid Ontario Health Insurance Plan number.

Of the 13.6% (179/1316) of patients who were eligible to participate, 46.9% (84/179) declined to participate, and 53.1% (95/179) consented to participate in the study. Table 4 lists the reasons for patients declining to participate.

The majority of participants enrolled in the study were recruited from the Mood and Anxiety Outpatient Clinic (n=92), and 3 were recruited from the Geriatric Psychiatry Program at the ROMHC. No participants were enrolled from the Youth Psychiatry Program. Table 5 describes the demographic characteristics of the participants. There were significantly more women in the intervention group and more men in the control group (X^2_1=6.6; P=.01).

Participants in the intervention group completed a mean number of 5.0 (SD 2.3) lessons in *The Journal* and 8.8 (SD 3.1) sessions with the coach. In the control group, 10% (5/48) of participants reported accessing *The Journal*. In the control group, 17% (8/48) of participants scored 9 or below on the PHQ-9 at baseline compared with 23% (11/47) of participants in the intervention group. Owing to gender imbalances between the groups, we also conducted a post hoc gender analysis of changes in PHQ-9 scores and health service use.

Figure 1. Consolidated Standards of Reporting Trials flow and attrition diagram. CONSORT: Consolidated Standards of Reporting Trials.

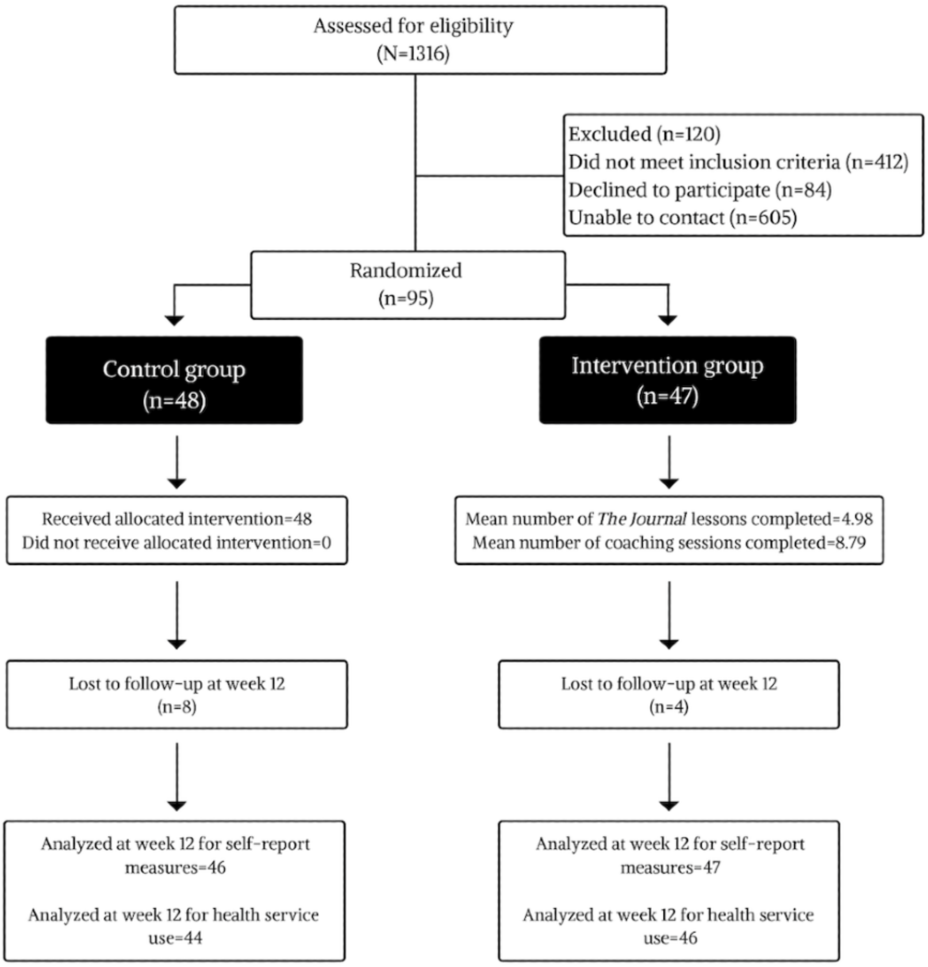


Table 4. Reasons for nonparticipation (n=84).

Reason for Nonparticipation	Value, n (%)
Did not attend baseline intake appointment	27 (32)
Did not have computer/internet at home—not interested in going to public library or community center	16 (19)
No reason provided	13 (15)
Interested in participating in a different study also recruiting from the mood and anxiety program	6 (7)
Too overwhelming	4 (5)
Would prefer to wait for appointment with psychiatrist	3 (4)
Not interested in participating in research at the Royal Ottawa Hospital	3 (4)
Did not feel that the study would benefit them	2 (4)
Not interested in weekly contact	2 (2)
No time	2 (2)
Family circumstances	1 (1)
Participating in another research study	1 (1)
Interested in medication change or recommendations	1 (1)
Migraines due to computer use	1 (1)
Moving out of province	1 (1)
Interested only in face-to-face therapy	1 (1)

Table 5. Sample demographic characteristics.

Demographic characteristic	Total (n=95)	Control group (n=48)	Intervention group (n=47)
Gender, n (%)[a]			
Male	28 (30)	20 (42)	8 (17)
Female	66 (70)	28 (58)	38 (83)[b]
Age (years), mean (SD)	44.2 (12.9)	44.8 (13.7)	43.5 (12.1)
Ethnicity, n (%)			
First Nations	1 (1)	0 (0)	1 (2)
Inuk	0 (0)	0 (0)	0 (0)
Métis	3 (3)	3 (6)	0 (0)
Asian	3 (3)	2 (4)	1 (2)
African, Caribbean, or Black	0 (0)	0 (0)	0 (0)
White	82 (87)	42 (88)	40 (87)
Other	5 (5)	1 (2)	4 (9)
Marital status, n (%)			
Single	33 (35)	11 (23)	22 (47)
Common law	6 (6)	4 (8)	2 (4)
Married	35 (37)	18 (38)	17 (36)
Separated	4 (4)	3 (6)	1 (2)
Divorced	16 (17)	11 (23)	5 (11)
Widowed	1 (1)	1 (2)	0 (0)
Education level, n (%)			
High School	10 (11)	7 (15)	3 (6)
College	45 (47)	23 (48)	22 (47)
University—undergraduate	26 (27)	10 (21)	16 (34)
University—master's	10 (11)	5 (10)	5 (11)
University—doctorate	4 (4)	3 (6)	1 (2)
Employment status, n (%)			
Full-time	23 (24)	11 (23)	12 (26)
Part-time	8 (8)	2 (4)	6 (13)
Short-term disability	5 (5)	2 (4)	3 (6)
Long-term disability	31 (33)	18 (38)	13 (28)
Self-employed	7 (7)	3 (6)	4 (8)
Retired	10 (11)	5 (10)	5 (11)
Unemployed	11 (12)	7 (15)	4 (8)

[a] n=1 transgender participant removed from the analysis.

[b] *P*=.01.

Primary Outcome

The mean PHQ-9 score was lower in the intervention group than in the control group at all study time points (Table 6). The proportions of missing values were comparable between the 2 groups at all assessment points, with the exception of week 12 (2/46 , 4% in the control group compared with 0/47, 0% in the intervention group). The Shapiro-Wilk test of normality revealed that all data were normally distributed with the exception of the control group at the week 12 study time point (W_{46}=0.95; *P*=.04).

At 12 weeks, the mean PHQ-9 score in the intervention group was 11.3 (SD 6.4) and was 12.4 (SD 6.4) in the control group (t_{86}=0.76; *P*=.45, 95% CI −1.9 to 3.8). To account for the non-normality in scores at week 12, mean differences (MDs) at this time were also assessed using a Mann-Whitney U test of significance, which confirmed that the difference between the two groups was not statistically significant (Mann-Whitney U statistic=862.0; Z=−0.85; *P*=.40). The mean change in PHQ-9 scores from baseline to week 12 was −3.6 (SD 6.6) in the intervention group and −3.1 (SD 6.2) in the control group, which was not statistically significant (t_{91}=−0.37; *P*=.72, 95% CI −3.1 to 2.2).

Excluding the 10% (5/48) of participants in the control group who also accessed *The Journal* did not result in any statistically significant differences in mean PHQ-9 scores compared with the intervention group. The mean score at week 12 for the control group excluding those who accessed *The Journal* (n=41) was 12.4 (SD 6.2) compared with 10.7 in the intervention group (t_{86}=0.76; *P*=.45, 95% CI −0.9 to 4.3).

Table 6. Mean and median scores on the Patient-Health Questionnaire-9 (n=95).

Study time point	Mean (SD)	Median	Missing values, n (%)	Mean difference (95% CI)	Independent samples *t* test (*df*)	*P* value
Baseline				**0.5 (−1.8 to 2.9)**	**0.45 (93)**	**.66**
Control (n=48)	15.4 (5.4)	16.5	0 (0)			
Intervention (n=47)	14.9 (6.0)	16.0	0 (0)			
Week 2				**1.1 (−1.3 to 3.5)**	**0.92 (89)**	**.36**
Control (n=46)	12.6 (5.5)	11.5	2 (4)			
Intervention (n=45)	11.5 (5.9)	12.0	2 (4)			
Week 6				**1.7 (−0.9 to 4.3)**	**1.29 (89)**	**.19**
Control (n=46)	12.4 (6.2)	12.0	2 (4)			
Intervention (n=45)	10.7 (6.1)	10.0	2 (4)			
Week 12				**1.0 (−1.9 to 3.8)**	**0.76 (89)**	**.45**
Control (n=46)	12.4 (6.4)	11.5	2 (4)			
Intervention (n=47)	11.3 (6.4)	10	0 (0)			

The repeated measures modeling found that scores on the PHQ-9 differed significantly by study time point, irrespective of group allocation ($F_{2.6,222.8}$=11.59, *P*<.001). However, this relationship was not sustained once treatment group was taken into account ($F_{1,87}$=1.46, *P*=.23). Similarly, modeling found that scores on the PHQ-9 were not significantly associated with gender ($F_{1,87}$=0.33, *P*=.57) or the interaction between treatment group and PHQ-9 scores ($F_{1,87}$=0.70; *P*=.53). However, results were significantly associated with age of participants ($F_{1,87}$=5.97, *P*=.02, η^2=0.06).

Secondary Outcomes

Health-Related Quality of Life

After 12 weeks of treatment, participants in the intervention group reported higher health-related quality of life, with mean index scores on the EQ-5D-3L of 0.7 (SD 0.7) for the intervention group and 0.6 (SD 0.5) for the control group (t_{86}=−2.30; *P*=.02, CI 95% −0.2 to −0.1; Multimedia Appendix 3). Similarly, at 12 weeks, the mean EQ-5D-VAS score was significantly greater in the intervention group (mean 66.8, SD 18.0) than in the control group (mean 55.9, SD 19.2; t_{84}=−2.73; *P*=.01).

Changes in Suicidality

At 12 weeks, of the 47 participants in the intervention group, 66% (31/47) reported no suicidality, 28% (13/47) reported suicidality several days in the preceding 12 weeks, 2% (1/47) reported suicidality nearly half the days, and 4% (2/47) reported suicidality nearly every day. In comparison, of the 46 participants in the control group, 69% (33/46) reported no suicidality, 17% (8/46) reported suicidality several days, 6% (3/46) reported suicidality more than half the days, and 4% (2/46) reported suicidality nearly every day (Table 7). These differences in suicidality were not statistically significant (X^2_3=2.2; *P*=.52).

Table 7. Mean and median scores on the Patient-Health Questionnaire -9, Question 9 (n=95).

Study time point	Control (n=48), n (%)	Intervention (n=47), n (%)	Chi-square (*df*)	*P* value
Baseline			**0.44 (3)**	**.93**
Not at all	24 (50)	23 (49)		
Several days	16 (33)	18 (38)		
More than half the days	5 (11)	4 (9)		
Nearly everyday	3 (6)	2 (4)		
Total	48 (100.0)	47 (100.0)		
Week 2			**1.37 (3)**	**.71**
Not at all	30 (63)	34 (72)		
Several days	12 (25)	8 (17)		
More than half the days	2 (4)	2 (4)		
Nearly everyday	2 (4)	1 (2)		
Total	46 (96)	45 (95)		
Week 6			**4.70 (3)**	**.20**
Not at all	27 (56)	32 (68)		
Several days	11 (23)	11 (23)		
More than half the days	5 (10)	2 (4)		
Nearly everyday	3 (6)	0 (0.0)		
Total	46 (95)	45 (95)		
Week 12			**2.24 (3)**	**.52**
Not at all	33 (69)	31 (66)		
Several days	8 (17)	13 (28)		
More than half the days	3 (6)	1 (2)		
Nearly everyday	2 (4)	2 (4)		
Total	46 (96)	47 (100.0)		

Hospital Service Use

Participants in both groups received similar levels of face-to-face follow-up care, with 25.0% of participants in the intervention group receiving any follow-up in the next 12 months at the hospital after their initial outpatient appointment with a psychiatrist compared with 21.3% of the control group (X^2_1=0.7; *P*=.80; Table 8). Participants in the intervention group attended a mean number of 4.8 (SD 8.5) face-to-face follow-up appointments compared with 3.6 (SD 5.6) in the control group (Mann-Whitney U statistic=255.50; Z=−0.11; *P*=.91).

Table 8. Hospital service use.

Health service use indicator	Control	Intervention	Tests of association	*P* value
Number of days from referral to first appointment[a], mean (SD)	213.6 (54.6)	219.3 (57.0)	t_{69}=−0.43 *P*	.67
Number of people who received outpatient follow-up by a nonpsychiatrist after initial assessment by a psychiatrist, n (%)	13/44 (30)	12/46 (26)	X^2_1=0.1	.71
Number of people who received outpatient follow-up by a psychiatrist after their initial consultation, n (%)	20/44 (42)	18/46 (38)	X^2_1=0.3	.54
Number of outpatient follow-up appointments with a psychiatrist in the year after the initial consultation[b], mean (SD)	2.3 (3.1)	2.5 (4.6)	t_{88}=−0.30	.76
Number of outpatient follow-up appointments with all disciplines in the year after initial consultation[c], mean (SD)	3.6 (5.6)	4.8 (8.5)	M-W[d] U=255.5, Z=−0.11	.91

[a] Control group n=34; Intervention group n=37.

[b] Control group n=34; Intervention group n=37.

[c] Control group n=34; Intervention group n=37.

[d] M-W: Mann-Whitney.

Process Evaluation

Implementation

Reach

As reported above, only 13.4% (95/711) of potentially eligible patients (eg, those who could be contacted and were on the waitlist for mental health treatment) agreed to participate in the trial. Of particular note is that 8.9% (16/179) eligible participants declined to participate because they did not have computer or internet access at home. This is comparable with 14% of the Canadian population [33]. Similarly, among participants who were randomized to one of the trial arms, 42% (40/95) were university-educated compared with 33.3% of residents of Ottawa [34].

Fidelity

Analysis of trial records demonstrated that the study intervention was implemented as intended in the study protocol. Participants in the intervention group completed a mean of 5.0 (SD 2.3) lessons in *The Journal*. Similarly, throughout the course of the study, participants completed a mean of 8.8 (SD 3.1) telephone coaching sessions, ranging from a mean of 9.6 to 61.6 min in duration, and a mean of 21.6 (SD 10.8) contacts with the coach (mean 13.1 (SD 4.0) telephone calls, 2.0 (SD 2.5) emails, and mean 25.7 (SD 15.4) text messages. All fidelity measures were significantly associated with PHQ-9 scores at 12 weeks, with the exception of the types of contact with the coach (Table 9). Total lessons completed in *The Journal*, total number of coaching sessions completed, and average length of coaching calls were all inversely related to PHQ-9 scores at 12 weeks, with those having completed more lessons in *The Journal*, a higher number of sessions with the coach and longer coaching calls reporting lower levels of depression after 12 weeks of treatment (Table 9).

Table 9. Relationship between fidelity measures and Patient-Health Questionnaire scores at 12 weeks (n=47).

Fidelity Measures	Mean (SD)	Median	Pearson’s correlation with PHQ-9[a] scores at week 12, (n=47)	
			r	*P* value
Total lessons completed in *The Journal*	5.0 (2.3)	6.0	−0.436	.002
Total sessions with the coach	8.8 (3.1)	10.0	−0.435	.002
Average length of coaching calls (min)	30.8 (12.9)	27.7	−0.360	.01
Total contacts with coach by type				
Telephone	13.1 (4.0)	13.0	0.061	.68
Email	2.0 (2.5)	1.0	−0.163	.27
Text message	25.7(15.4)	26.0	−0.073	.62

[a] PHQ-9: Patient Health Questionnaire.

Dose

For those participants in the intervention group who received a full *dose* of the study intervention, completing 6 or more lessons from *The Journal* (n=29), the mean PHQ-9 scores at 12 weeks was 9.5 (SD 5.7) compared with 14.2 (SD 6.6) among those who completed 5 lessons or less in *The Journal* and 12.4 (SD 6.4) in the control group ($F_{2,85}$=3.5; *P*=.04). Post hoc comparisons using the Tukey Honestly Significant Difference test indicated that the MD among those who completed at least 6 lessons in *The Journal* was significantly different from those who completed 5 or fewer lessons in *The Journal* (MD=−4.7, *P*=.04). However, the MDs between the control group and both experimental groups were not statistically significant ($MD_{\geq 6\ Lessons}$=−2.8; *P*=.15; $MD_{\leq 5\ Lessons=}$1.86; *P*=.55). Similarly, those who completed at least six or more lessons in *The Journal* were more likely to demonstrate a clinically significant reduction in symptoms, with 65.5% reporting a significant reduction in symptoms compared with 27.8% of those who completed less than 5 or fewer lessons and 37.0% of the control group (X^2_2=8.2; *P*=.02; Φ=0.30).

Similarly, by week 12 in the intervention group, 51% (24/47) of participants had a clinically important reduction in depressive symptoms (PHQ-9 score of 9 or less or a 50% or more improvement in scores) compared with 37% (17/46) of participants in the control group. However, this difference was not statistically significant (X^2_1=1.8; *P*=.17). Among participants in the intervention group who completed the full *dose* of the intervention by finishing 6 or more *Journal* sessions, 66% (19/29) had clinically important reductions in depression at 12 weeks compared with 37% (17/46) in the control group (X^2_1=6.3; *P*=.01; Φ=0.37).

Context and Mechanisms of Impact

A total of 15 qualitative interviews were conducted (10 intervention, 5 control) with 4 male (2 intervention group, 2 control group), 10 female (7 intervention group, 3 control group), and 1 transgender person (intervention group). Interviews took place at the ROMHC at a time that was mutually convenient for both participants and interviewers. Themes that arose from the interviewers were categorized in terms of whether there were positive facilitators of or negative barriers to engagement with the study intervention.

Facilitators to Engagement With The Journal

Participants identified Sir John Kirwan's relatability as a positive facilitator through his sharing of his personal experience with depression. For instance, a participant commented on the videos that begin each lesson in *The Journal*, explaining the following:

> *…I just like the way that they interacted with each other and how casual it was, so it didn't feel like I was doing homework, and it didn't feel like I was doing medical stuff…It was just, watching two people sitting on a bench in a park and they were talking, and I really liked how the narrator came right out and said, you know, when I was depressed, you know, I couldn't get out of bed, or I, I didn't want to take a shower, it was, like, wow somebody else feels the way I do. So, it helped me to feel like I wasn't all, all alone. And he was, he seemed to be very honest about his experience.* [P040]

Similarly, participants expressed an appreciation for the structure and layout of *The Journal*, specifically commenting on the usefulness of the Problem-Solving modules:

> *Interviewer: …what did you like about the problem-solving lesson?*
>
> *Participant: … it was in a new, new approaches. It gave me a new technique or I found a new technique in there in terms of outlining problems, and thinking about them and, that I hadn't.* [P095]

Barriers to Engagement With The Journal

Participants also reported frustrations with some aspects of *The Journal*, specifically relating to the technical issues they encountered. At the time of the trial, *The Journal* was not available on mobile devices, such as smartphones or tablets, which may have limited participants' ability to access it. Similarly, technical issues presented themselves within the web-based program as well. For instance, in each lesson, participants must set deadlines for tasks that they are required to complete and are unable to move forward in *The Journal* until these deadlines pass. Participants also identified issues with motivation and questioned whether *The Journal* would be most appropriate for patients who are only mildly depressed. For instance, one participant explained:

> *Participant: Yeah, lifestyle one…was little bit of a challenge.*
>
> *Interviewer: Yeah? In what way?*
>
> *Participant: … just trying to find motivation…it's just hard to change some things about lifestyle, but it came with time"* [P002]

Similarly, a participant in the control group commented on how some depressive symptoms, such as difficulty with concentration, may act as barriers to engagement with treatment:

> *Interviewer: … Were there any barriers or challenges to using the brochure?*
>
> *Participant: …not really because it's short and sweet…my only barrier was time, you know, time and, and energy. Because at the time, and it was almost an opportune time that I was doing this because I was low at the time… so when you are depressed the last thing you want to do it read, so you can't even. And that's why I was saying that if this was little bullet chunks, you know, you don't have to read it so much, cause I can't read a book when I'm down. Like I start to read a paragraph and I'm like, no I have to close it.* [P014]

Coach as a Mechanism of Change

Numerous participants in the intervention group also identified the coach as key to their success with the program, highlighting the importance of accountability:

> *Interviewer: …what would you say made the biggest difference to your participation in the study? ...was it the journal, was it the coach?*
>
> *Participant: It was the coach, oh my god. I would not have stayed with that journal unless [SL] was calling me… cause then I wouldn't really have anybody to be like, hey did you do that thing?* [P038]

Similarly, another participant questioned whether they would have been as successful in completing the program had they not had someone following-up on their goals:

> *…if it was by myself, I might have just kept postponing it or not doing it or, so I found it quite helpful. She's a pretty great coach.* [P068]

Similarly, another participant highlighted the importance of continuity between the lessons in *The Journal* and the content of the coaching sessions:

> *I like the conversations the following week with the homework that we did. I liked that it was continuous.* [P069]

Accountability also emerged as important for participants in the control group. For instance, 1 participant who accessed *The Journal* expressed a lack of follow-up by a third party as a barrier to completing the program, suggesting that:

> *…you guys can follow-up and see if they're.. If people are using it or not…just to say, somebody is following me, I should probably do a little bit more of this...* [P066]

This was echoed by another participant in the control group who struggled with the limited follow-up they received from the research team:

> *taking not even five minutes to say "how are you doing right now?"… So in a way it gauges their, their level of health, of mental health at the time, and if they do appear to need resources then [the Research Assistant] could say "look I'm not a doctor, I'm not a psychiatrist, but do you remember the flyer that I gave you, on it there was this thing that might be helpful for you right now, you might want to give it a thought"… so, that's just like extra, like value added for resourcefulness.* [P014]

Discussion

Principal Findings

Among patients on a waiting list referred to secondary care for depressive symptoms, a trial of usual care plus coach-facilitated web-based therapy compared with usual care plus information-only waitlist control group found no statistically significant differences in mean depression scores after 12 weeks. However, participants who completed 6 or more lessons in *The Journal* reported significantly greater reductions in symptoms compared with controls. Participants in the intervention group also reported better quality of life after 12 weeks of treatment. There was no difference in service use between the two groups. During the qualitative interviews conducted as part of the process evaluation, participants identified the role of the coach as a major factor in their completion of the web-based therapy program.

Strengths and Limitations

The major strength of this study is that it is the first study of coach-assisted web-based therapy in a secondary care setting in Canada. This study demonstrated that it is possible to incorporate the use of a coach within this clinical setting. However, due to limitations imposed by institutional policies, study staff were restricted in their ability to recruit participants, with only 12% (95/784) of potentially eligible patients agreeing to participate in the trial. It is reasonable to expect greater reach if this intervention was rolled out by a clinical service that could approach people directly at the time of referral.

There was also a possibility of contamination between the arms of the trial. Given that *The Journal* is a publicly accessible website, any participant was able to access it, and there was little that could be done by the study team to prevent this. However, our experience is that patients in secondary care do not widely use *The Journal*, with only 10% (5/48) participants in the control group reporting use of *The Journal* during the treatment period. Furthermore, contamination is likely to bias the study to showing no difference (as the control group could use *The Journal* unguided), so any differences that are found are likely to be more *believable*. Finally, the ability to generalize study findings to the larger Canadian population may be limited as a result of our highly educated population.

Some participants who were interviewed as part of the process evaluation expressed issues with recall given that the interviews took place 6 months after study completion. It is possible that interviews conducted closer to study completion would have yielded richer qualitative data.

Comparison With Prior Work

A key contribution of this study is that participant engagement with and adherence to the intervention was much higher than that reported in previous work, with 62% (29/47) of participants in the intervention group receiving a full *dose* of treatment, completing an average of 5 out of 6 lessons in *The Journal* and 8.8 telephone coaching sessions with the coach. Comparatively, in the REEACT trial, participants only completed a median of 2 sessions of web-based therapy. Similarly, in the study by Kenter et al [21], only 36% of participants received the required 4 to 5 sessions of internet-based psychotherapy. This was found in spite of difficulties with ease of access, as *The Journal* was not available on mobile or tablet devices at the time of the study. This trial also highlights the importance of the uptake of the intervention to clinical outcomes, with participants who completed at least 6 lessons in *The Journal* experiencing a significantly greater reduction in depression symptoms than those who did not.

The previous trial of a guided version of *The Journal* in New Zealand secondary care showed no differences in clinical outcomes or service use. However, the patients in the New Zealand study had higher baseline depression scores, a mean PHQ-9 of 17.0, and were less engaged with the study intervention, completing only a mean of 2.5 lessons in *The Journal*. The magnitude of change in PHQ-9 scores was also different, with an intervention group change in PHQ-9 scores of 3.5 in Canada compared with 9.4 in New Zealand. This presumably reflects differences in usual care, with participants in New Zealand receiving mental health care, whereas Canadian participants received care from their primary care physicians while on a waiting list.

Clinical Implications

Web-based therapies are often promoted as a way to address long waiting lists for mental health care services. However, in this study, even though participants in the intervention group with more exposure to the study intervention were more likely to experience a significant reduction in symptoms, this had little impact on subsequent service use. Web-based therapies are part of a complex socio-technological system and, as such, cannot exist in a vacuum. To achieve improvement in patient outcomes, they must be integrated into a larger system of care. This could be achieved with a more explicit stepped care system supported by web-based therapies, with a more flexible response from providers based on patient need.

Unanswered Questions and Future Research

This study provides limited support for the potential use of web-based therapies within a stepped-care approach to the treatment of depression. However, an RCT is needed to determine the effectiveness of such an approach. The impact of providing digital services to those in greatest need, who are also the least likely to have access to high-speed connections, also needs to be taken into account. Future research on internet-based psychotherapy for depression needs to include the system of care in which it is used. This can be done through implementation science tools, which not only evaluate the effectiveness of web-based therapies but also the factors central to their uptake, such as reach, adoption, and sustainability.

Conclusions

The results of this study demonstrated that the use of guided web-based therapy for the treatment of depression is not more effective than information-only waitlist control. However, it showed that coach-guided web-based therapy has the potential to increase adherence and engagement with depression treatment protocols. Greater adherence resulted in greater effectiveness. More research is needed on the human component of coaching in conjunction with web-based therapy to examine what factors lead to greater adherence. Researchers also need to consider when and how web-based therapies should be integrated into existing clinical pathways.

Acknowledgments

This work was supported by a Canadian Institutes for Health Research Catalyst grant (funding reference number: 134160) and the Royal’s University Medical Research Fund. The funders had no role in the review or approval of the manuscript for publication.

The research team would like to recognize the following psychiatric programs at the ROMHC: Mood and Anxiety, Geriatric Psychiatry, and Youth Psychiatry, as well as the service users who participated in our study.

Authors' Contributions

SH conceived of the study and was the grant holder. SM took the lead in drafting of the manuscript, completed participant recruitment, and data collection activities. DJC conducted all statistical analyses. SL completed participant follow-up and data collection activities. JK and ZS completed the process evaluation of qualitative interviews. JK and KG conducted qualitative analysis of the process. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT EHEALTH Checklist V1.6.
[PDF File (Adobe PDF File), 20847 KB - jmir_v22i6e15001_app1.pdf]

Multimedia Appendix 2
Supplemental file – Screenshots from The Journal.
[DOCX File , 489 KB - jmir_v22i6e15001_app2.docx]

Multimedia Appendix 3
Mean and median scores on the EuroQol 5 dimension (3 levels) visual analog scale (n=94).
[DOCX File , 14 KB - jmir_v22i6e15001_app3.docx]

References

1. Wang PS, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 2007 Sep 8;370(9590):841-850 [FREE Full text] [doi: 10.1016/S0140-6736(07)61414-7] [Medline: 17826169]
2. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
3. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016 Jul;3(7):646-659. [doi: 10.1016/S2215-0366(16)30030-X] [Medline: 27289303]

4. NICE. National Institute for Health and Care Excellence. London, UK: National Institute for Health and Care Excellence; 2009 Oct 28. Depression in Adults: Recognition and Management Management. Clinical Guideline URL: https://www.nice.org.uk/guidance/cg90/resources/depression-in-adults-recognition-and-management-pdf-975742636741 [accessed 2019-10-25]
5. Stéphan F, Le Galudec M, Tritschler-Le Maitre M, Berrouiguet S, Walter M. Review of various cognitive behavioral therapy programs on the Internet for the treatment of depression. Article in French. Revue des différents programmes de thérapies cognitivo-comportementales sur Internet dans le traitement de la dépression. J Thérapie Comport Cogn 2014 Dec;24(4):138-143. [doi: 10.1016/j.jtcc.2014.05.002]
6. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
7. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008 Mar 25;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
8. Donker T, Bennett K, Bennett A, Mackinnon A, van Straten A, Cuijpers P, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. J Med Internet Res 2013 May 13;15(5):e82 [FREE Full text] [doi: 10.2196/jmir.2307] [Medline: 23669884]
9. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
10. Lewis CE, Farewell D, Groves V, Kitchiner NJ, Roberts NP, Vick T, et al. Internet-based guided self-help for posttraumatic stress disorder (PTSD): randomized controlled trial. Depress Anxiety 2017 Jun;34(6):555-565. [doi: 10.1002/da.22645] [Medline: 28557299]
11. Dear BF, Fogliati VJ, Fogliati R, Johnson B, Boyle O, Karin E, et al. Treating anxiety and depression in young adults: a randomised controlled trial comparing clinician-guided versus self-guided Internet-delivered cognitive behavioural therapy. Aust N Z J Psychiatry 2018 Jul;52(7):668-679. [doi: 10.1177/0004867417738055] [Medline: 29064283]
12. Topooco N, Berg M, Johansson S, Liljethörn L, Radvogin E, Vlaescu G, et al. Chat- and internet-based cognitive-behavioural therapy in treatment of adolescent depression: randomised controlled trial. BJPsych Open 2018 Jul;4(4):199-207 [FREE Full text] [doi: 10.1192/bjo.2018.18] [Medline: 29988969]
13. Rollman BL, Belnap BH, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: A randomized clinical trial. JAMA Psychiatry 2018 Jan 1;75(1):56-64 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
14. Arjadi R, Nauta MH, Scholte WF, Hollon SD, Chowdhary N, Suryani AO, et al. Internet-based behavioural activation with lay counsellor support versus online minimal psychoeducation without support for treatment of depression: a randomised controlled trial in Indonesia. Lancet Psychiatry 2018 Sep;5(9):707-716. [doi: 10.1016/S2215-0366(18)30223-2] [Medline: 30006262]
15. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs technician assistance. PLoS One 2010 Jun 8;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
16. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
17. Thompson SC, Schlehofer MM. The many sides of control motivation: motives for high, low, and illusory control. In: Shah JY, Gardner WL, editors. Handbook of Motivation Science. New York, NY: The Guilford Press; 2008:41-56.
18. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 3;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
19. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
20. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
21. Kenter RM, Cuijpers P, Beekman A, van Straten A. Effectiveness of a web-based guided self-help intervention for outpatients with a depressive disorder: Short-term results from a randomized controlled trial. J Med Internet Res 2016 Mar 31;18(3):e80 [FREE Full text] [doi: 10.2196/jmir.4861] [Medline: 27032449]

22. Hatcher S, Whittaker R, Patton M, Miles WS, Ralph N, Kercher K, et al. Web-based therapy plus support by a coach in depressed patients referred to secondary mental health care: randomized controlled trial. JMIR Ment Health 2018 Jan 23;5(1):e5 [FREE Full text] [doi: 10.2196/mental.8510] [Medline: 29362207]
23. Depression and Anxiety. Wellington, NZ: New Zealand Health Promotion Agency; 2019. The Journal: Learn How to Feel Better with The Journal URL: https://depression.org.nz/get-better/the-journal/ [accessed 2019-10-25]
24. Hatcher S, Sharon C, Parag V, Collins N. Problem-solving therapy for people who present to hospital with self-harm: Zelen randomised controlled trial. Br J Psychiatry 2011 Oct;199(4):310-316. [doi: 10.1192/bjp.bp.110.090126] [Medline: 21816868]
25. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29(5):388-395. [doi: 10.1016/j.genhosppsych.2007.06.004] [Medline: 17888804]
26. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep 1;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
27. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
28. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
29. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
30. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 1997 May;36(5):551-559. [doi: 10.1093/rheumatology/36.5.551] [Medline: 9189057]
31. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. J Affect Disord 2010 Dec;127(1-3):122-129. [doi: 10.1016/j.jad.2010.04.030] [Medline: 20569992]
32. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
33. Canadian Internet Registration Authority. Ottawa, ON: Canadian Internet Registration Authority; 2019. Canada's Internet Factbook URL: https://cira.ca/factbook/canada%E2%80%99s-internet-factbook-2018 [accessed 2019-07-03]
34. Statistics Canada. Ottawa, ON: Statistics Canada; 2017. Census Profile, 2016 Census URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=505&Geo2=PR&Code2=35&SearchText=Ottawa+-+Gatineau&SearchType=Begins&SearchPR=01&B1=Education&TABID=1&type=0 [accessed 2019-07-03]

Abbreviations

CBT: Cognitive behavioral therapy
EMR: electronic medical record
EQ-5D-3L: EuroQol 5 Dimensions (3 levels)
EQ-5D-VAS: EuroQol 5 dimensions visual analogue scale
MD: mean difference
NICE: National Institute for Health and Care Excellence
PHQ-9 Q9: Patient Health Questionnaire question 9
PHQ-9: Patient Health Questionnaire
RCT: randomized controlled trial.
REEACT: Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy
ROMHC: Royal Ottawa Mental Health Centre.

Edited by G Eysenbach; submitted 05.07.19; peer-reviewed by K Stasiak, D Leightley, L Saslow; comments to author 30.08.19; revised version received 25.10.19; accepted 17.12.19; published 09.06.20.

Please cite as:
MacLean S, Corsi DJ, Litchfield S, Kucharski J, Genise K, Selaman Z, Testa V, Hatcher S
Coach-Facilitated Web-Based Therapy Compared With Information About Web-Based Resources in Patients Referred to Secondary Mental Health Care for Depression: Randomized Controlled Trial
J Med Internet Res 2020;22(6):e15001
URL: https://www.jmir.org/2020/6/e15001
doi:10.2196/15001
PMID:32515740

Original Paper

User-Centered Design of a Mobile Health Intervention to Enhance Exacerbation-Related Self-Management in Patients With Chronic Obstructive Pulmonary Disease (Copilot): Mixed Methods Study

Yvonne J G Korpershoek[1], MSc; Sander Hermsen[2], PhD; Lisette Schoonhoven[3], Prof Dr; Marieke J Schuurmans[4], Prof Dr; Jaap C A Trappenburg[3], PhD

[1]Research Group Chronic Illnesses, University of Applied Sciences Utrecht, Utrecht, Netherlands
[2]OnePlanet Research Center, imec NL, Wageningen, Netherlands
[3]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands
[4]Education Center, UMC Utrecht Academy, University Medical Center Utrecht, Utrecht, Netherlands

Corresponding Author:
Yvonne J G Korpershoek, MSc
Research Group Chronic Illnesses
University of Applied Sciences Utrecht
Heidelberglaan 7
Utrecht, 3584 CS
Netherlands
Phone: 31 6 38763949
Email: yvonne.korpershoek@hu.nl

Abstract

Background: Adequate self-management skills are of great importance for patients with chronic obstructive pulmonary disease (COPD) to reduce the impact of COPD exacerbations. Using mobile health (mHealth) to support exacerbation-related self-management could be promising in engaging patients in their own health and changing health behaviors. However, there is limited knowledge on how to design mHealth interventions that are effective, meet the needs of end users, and are perceived as useful. By following an iterative user-centered design (UCD) process, an evidence-driven and usable mHealth intervention was developed to enhance exacerbation-related self-management in patients with COPD.

Objective: This study aimed to describe in detail the full UCD and development process of an evidence-driven and usable mHealth intervention to enhance exacerbation-related self-management in patients with COPD.

Methods: The UCD process consisted of four iterative phases: (1) background analysis and design conceptualization, (2) alpha usability testing, (3) iterative software development, and (4) field usability testing. Patients with COPD, health care providers, COPD experts, designers, software developers, and a behavioral scientist were involved throughout the design and development process. The intervention was developed using the behavior change wheel (BCW), a theoretically based approach for designing behavior change interventions, and logic modeling was used to map out the potential working mechanism of the intervention. Furthermore, the principles of design thinking were used for the creative design of the intervention. Qualitative and quantitative research methods were used throughout the design and development process.

Results: The background analysis and design conceptualization phase resulted in final guiding principles for the intervention, a logic model to underpin the working mechanism of the intervention, and design requirements. Usability requirements were obtained from the usability testing phases. The iterative software development resulted in an evidence-driven and usable mHealth intervention—Copilot, a mobile app consisting of a symptom-monitoring module, and a personalized COPD action plan.

Conclusions: By following a UCD process, an mHealth intervention was developed that meets the needs and preferences of patients with COPD, is likely to be used by patients with COPD, and has a high potential to be effective in reducing exacerbation impact. This extensive report of the intervention development process contributes to more transparency in the development of complex interventions in health care and can be used by researchers and designers as guidance for the development of future mHealth interventions.

(J Med Internet Res 2020;22(6):e15449) doi:10.2196/15449

KEYWORDS

mobile health; mHealth; user-centered design; behavior change; COPD; exacerbation; self-management; self-care; mobile phone

Introduction

Background

Chronic obstructive pulmonary disease (COPD) is a highly prevalent chronic disease and is predicted to be the third leading cause of mortality worldwide in 2030 [1,2]. Exacerbations are important events in the course of COPD, as they accelerate the decline in lung function [3], negatively affect the quality of life [4,5], and lead to increased mortality and high socioeconomic costs [6,7]. An exacerbation is defined as *a sustained worsening of patients'respiratory symptoms, which are beyond normal day-to-day variability and may warrant medical treatment* [8]. The absence of an adequate imminent exacerbation marker requires a focus on supporting patients with COPD in developing self-management skills to reduce the impact of exacerbations [9]. Self-management is defined as *an individual's ability to detect and manage symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent in living with a chronic condition* [10].

Recent interventions focusing on exacerbation-related self-management (including the use of action plans) have shown positive outcomes on quality of life and hospital admissions [11,12]. However, there is still a substantial proportion of patients with COPD who barely benefit from these kinds of interventions [11-13]. This might be explained by the *one-size-fits-all* and static approach regarding design, intensity, and mode of delivery without a focus on individual exacerbation patterns and actions. Moreover, recent interventions have a strict focus on exacerbation detection and taking action and the use is suboptimal [11,14,15]. To further reduce the impact of exacerbations, more comprehensive, dynamic, and individualized strategies are needed to improve the full spectrum of exacerbation-related self-management behavior that meet patients' needs, perceptions, and capabilities [12,16].

Mobile health (mHealth) is considered promising in engaging patients in their own health and changing health behaviors [17,18]. The rapidly evolving nature and increased uptake of mHealth are bound to influence the accessibility and the way self-management support will be provided in the future, also in patients with COPD [19-21]. Recent studies suggest that mHealth interventions focusing on COPD self-management lead to positive outcomes, although no firm conclusions could be drawn because of poor quality and heterogeneity among the studies [19,20]. Nonetheless, the use of mHealth creates opportunities to strongly individualize interventions and to provide more dynamic and intensive therapeutic stimuli that fit with real-time health status and individual exacerbation patterns. As a result, mHealth can reach patients at the right moment and can provide tailored support anytime and anywhere, which could stimulate the development of effective self-management skills and change health behaviors.

To date, there is limited knowledge on how to design mHealth interventions that are effective, meet the needs of intended end users, and are perceived as useful [17,22]. Designing mHealth interventions to change health behaviors is complex and needs theoretical grounding to increase the design's efficacy. In current thinking about the development of behavior change interventions, the importance of theory is clear [23-26], but the way in which theory should be incorporated in the design process is not [24,27,28]. Furthermore, specific steps in the development of evidence- and theory-driven interventions that involve the end users are rarely described transparently in literature [22,29].

Objectives

During a 4-year period, our research team has developed an evidence-driven and usable mHealth intervention to enhance exacerbation-related self-management in patients with COPD. By following an iterative user-centered design (UCD) process, several studies were performed to increase the likelihood of developing an mHealth intervention that is effective, fits with patients' needs and preferences, and can be successfully implemented in routine COPD care. Some of these studies have recently been published [15,30,31]. This paper underpins the design and working mechanism of this COPD-specific mHealth intervention and offers a novel and potentially effective method to use evidence and theory to inform the design of mHealth interventions in general.

The aim of this paper was to describe in detail the full UCD and development process of an evidence-driven and usable mHealth intervention to enhance exacerbation-related self-management in patients with COPD, including the design, iterative software development, and usability testing.

Methods

User-Centered Design Process

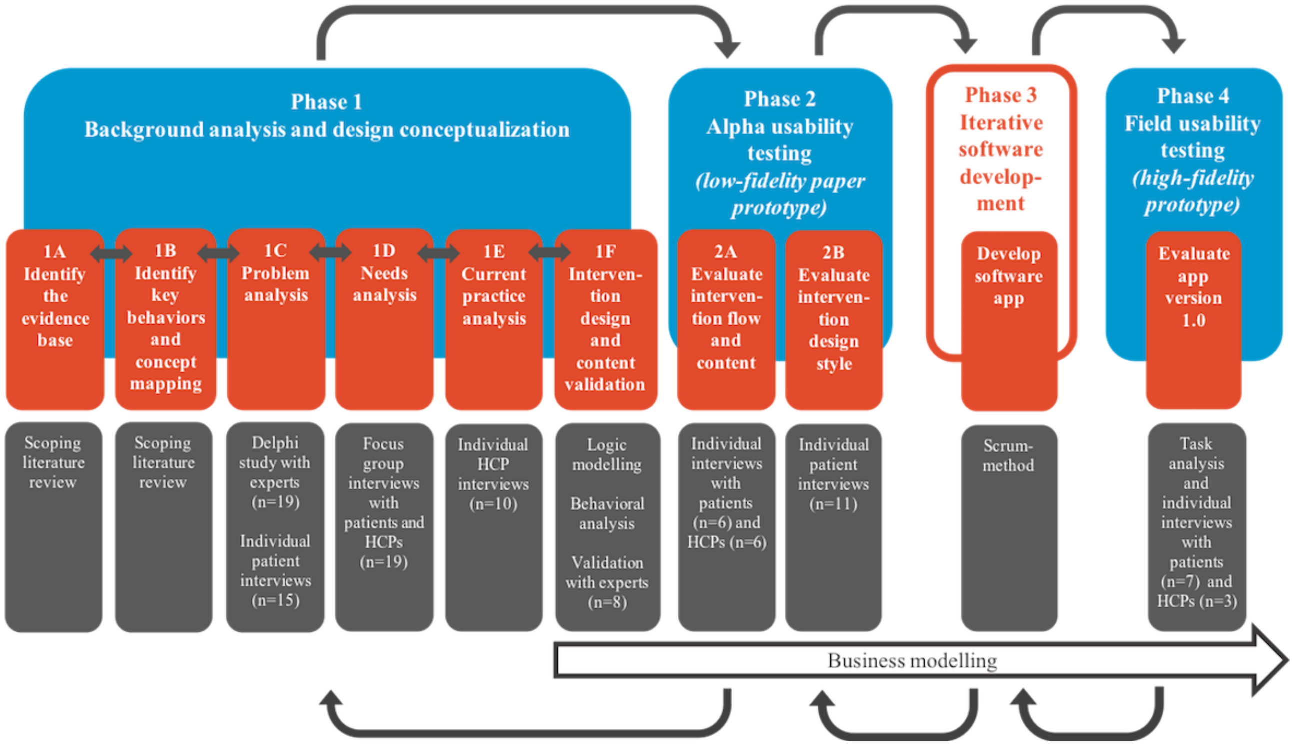
Guiding principles for the mHealth intervention were formulated by the research team at an early stage to provide a framework for making decisions during intervention development (Textbox 1) [32]. The guiding principles were based on recent evidence regarding COPD self-management and were progressively refined as the intervention development proceeded based on outcomes of specific development steps that we described in this paper. The mHealth intervention was developed by following a UCD process involving patients with COPD, health care providers (HCPs), COPD experts, designers, software developers, and a behavioral scientist. The UCD was based on the methodology as described by Johnston et al [33] consisting of four iterative phases: (1) background analysis and design conceptualization, (2) alpha usability testing, (3) iterative software development, and (4) field usability testing (Figure 1) [33]. Johnston et al [33] provide limited guidance on the specific steps needed to develop an effective mHealth intervention that meets patients' needs and preferences and fits with current COPD care. Therefore, we extended the first phase of the UCD with subphases based on a comprehensive approach that combines elements of the Medical Research Council (MRC) framework development phase with elements of existing development models (Figure 1) [34]. The MRC framework is a well-known and often used framework for the development and evaluation of complex interventions in health care with a specific focus on developing theory- and evidence-driven interventions. The whole design and development process was carried out between 2015 and 2019. The methods of each phase are chronologically described in the following paragraphs. The results of each phase are detailed in the Results section.

Textbox 1. Guiding principles for a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease.

The mobile health intervention should:

- meet individual patient needs, perceptions, and preferences regarding exacerbation-related self-management;
- synchronize with current health status and anticipate on the heterogeneity of exacerbations in and between patients;
- focus on target behaviors in the full spectrum of exacerbation-related self-management;
- include a chronic obstructive pulmonary disease (COPD) action plan along with ongoing self-management support;
- focus on the continuous development of self-management skills and behavior change;
- stimulate proactive self-monitoring;
- be safe, literacy-sensitive, and patient-friendly;
- be feasible in current Dutch COPD care; and
- meet the conceptual definition of a COPD self-management intervention: *A COPD self-management intervention should be structured but personalized and often multicomponent, with goals of motivating, engaging and supporting the patients to positively adapt their health behavior(s) and develop skills to better manage their disease* [16].

Figure 1. User-centered design for the development of the mobile health intervention. HCP: health care provider.



Phase 1: Background Analysis and Design Conceptualization

The aim of the first phase was to identify the evidence base and to achieve a theoretical understanding of the underlying process of change for the intervention [35].

1A: Identify the Evidence Base

In phase 1A, a scoping literature review was performed in Medical Literature Analysis and Retrieval System Online (MEDLINE) to explore current systematic reviews on interventions that focus on enhancing exacerbation-related self-management in patients with COPD, including mHealth interventions, and to identify potential effective intervention components (Figure 1; phase 1A). Literature review on interventions was an ongoing process during the whole intervention development process, to stay up to date on developments about (mHealth) interventions focusing on exacerbation-related self-management.

1B: Identify Key Behaviors and Concept Mapping

In phase 1B, a scoping literature review was performed in MEDLINE to specify symptom fluctuation phases during the course of COPD and to identify relevant self-management behaviors that can reduce exacerbation impact (Figure 1; phase 1B). Two researchers (YK and JT) developed a conceptual model of patients' fluctuations in symptoms during the course of COPD. Then, an initial set of relevant self-management behaviors was generated and added to the conceptual model. The methods of the scoping review and stepwise development of the conceptual model are published elsewhere [15].

1C: Problem Analysis

A problem analysis was included to provide insight into the problems experienced by patients and identified by experts to determine the intervention targets and to set boundaries of the intervention (Figure 1; phase 1C). A two-round Delphi study with 19 international respiratory experts (medical doctors and key researchers in the field of COPD) was performed. In this study, insight into expert opinion was provided on the most relevant set of self-management behaviors that have the potential to maximally reduce the impact of exacerbations and is feasible to target and influence before, during, and after an exacerbation. The methodology is described in depth in the publication of this study [15]. Furthermore, a grounded theory study using individual in-depth interviews with patients with COPD (n=15) was performed [36]. In this study, patient perceptions, capabilities, and needs with regard to exacerbation-related self-management were explored to identify and explain the underlying process of exacerbation-related self-management behavior in patients with COPD. The methodology is described in depth elsewhere [30].

1D: Needs Analysis

Patients' needs regarding exacerbation-related self-management were partially identified in phase 1C because these needs flowed naturally from the problems perceived by patients [37]. An additional needs analysis was performed to further investigate specific needs and explicit requests for care with regard to using mHealth for self-management (Figure 1; phase 1D) [37]. To develop an mHealth intervention with optimal usability and feasibility, a deep and early understanding of both patients' and HCPs' perspectives was considered to be important [38]. Therefore, a qualitative study using focus group interviews with both patients with COPD (n=13) and HCPs (n=6) was performed to (1) explore their willingness to use mHealth for self-management of exacerbations, (2) identify potential benefits and barriers of using mHealth, and (3) explore needs and preferences regarding the content of an mHealth intervention [39]. The methods of this step are further described in the paper of this study [31].

1E: Current Practice Analysis

An analysis of COPD guidelines and current practice was performed to gain insight into current exacerbation-related self-management support and to explore the added value of the intended intervention compared with regular care (Figure 1; phase 1E). Individual semistructured interviews with HCPs (n=10) were performed to identify HCPs' perspectives with regard to care provided and their role in providing self-management support. Purposive sampling was performed in primary and secondary care settings. The following topics were discussed: current interventions to support exacerbation-related self-management, HCP experiences with providing self-management support, perceptions toward HCPs' roles and responsibilities, barriers in providing self-management support, and the potential to use mHealth for self-management support. All interviews were audiotaped, transcribed verbatim, and analyzed by open, axial, and selective coding [40].

1F: Intervention Design

The aim of the intervention design phase (Figure 1; phase 1F) was twofold: (1) to map out the potential working mechanisms triggered by the intervention and (2) to develop the flow and content of the intervention. During this phase, a decision was made on the target behaviors of the intervention. The behavior change wheel (BCW) method was used to analyze the target behaviors and to design intervention components [41]. First, based on the literature, behavioral analysis was performed by two researchers (YK and SH) to identify what needs to change in patients' capability, opportunity, and motivation to improve each target behavior (capability, opportunity, and motivation model of behavior [COM-B] analysis) [41]. Second, the theoretical domains framework (TDF) was used to elaborate on the behavioral analysis by mapping the 14 domains of the framework onto the capability, opportunity, and motivation components of COM-B [41]. Third, potentially relevant intervention functions and behavior change techniques (BCTs), matching users and context, were selected using criteria provided by the BCW [41] (see Multimedia Appendix 1). Logic modeling was used to map out the potential working mechanism of the intervention by detailing all evidence and assumptions underpinning the pathway from the intervention to the long-term impact on outcomes [42,43]. The logic model starts with the target behaviors and details what needs to change in behavior (TDF), by which intervention functions and BCTs, and through which specific intervention components, including factors that could influence the working mechanism, and results in short- and long-term outcomes. The logic model components were based on the evidence gained from all previous phases, and consensus on the components was reached during research group meetings. On the basis of this model, design requirements were formulated.

Furthermore, creative ideas with regard to the intervention design were explored using methods derived from design thinking [44]. In a *pressure cooker session* with three independent creative designers, initial ideas on the design were presented by focusing on potential techniques to change health behaviors and to enhance engagement with mHealth. After this session, collaboration with a creative design agency (Panton BV, Deventer, the Netherlands) specifically focusing on health care solutions was initiated. By following an iterative design process, the flow and content of the intervention were designed, and various design styles were developed using low-fidelity prototypes—paper prototypes that visualize design solutions. In the early stages of digital user interface design, such low-fidelity paper prototypes are often used to determine requirements for the architecture and functionalities of the specific intervention to be designed [45]. The paper prototypes were tested in phase 2 of the UCD.

Moreover, the content of a symptom-monitoring module was developed during this stage. The module aimed to determine the individual COPD patient's normal day-to-day variability in symptoms to be able to set the patient's normal symptom pattern. The content validity of the module was evaluated by experts in the field of COPD (n=8) according to the Lynn method [46]. Each symptom was rated on *relevance* and *linguistics* by answering four questions. All questions about *relevance* were rated on a 4-point Likert scale (1=not relevant, 4=relevant). *Linguistics* was determined by if the interpretation was clear (*yes* or *no*). The item-content validity index (I-CVI) was calculated for each *relevance* question to determine the number of experts judging the content as valid (I-CVI>0.78=relevant). Subsequently, the scale-content validity index (S-CVI) was calculated to determine the relevance of the whole symptom-monitoring module (S-CVI>0.90=excellent) [46]. *Linguistics* was considered to be clear when at least 75% of the expert panel rated clearness of interpretation as a *yes*. A more in-depth description of the development and content validity assessment of the symptom-monitoring module is given in Multimedia Appendix 2.

Phase 2: Alpha Usability Testing

In the second phase, alpha usability tests were performed by investigating patient and HCP responses to low-fidelity paper prototypes of the intervention in two steps: (1) evaluating the intervention flow and content and (2) evaluating intervention design styles [33]. At each phase of usability testing, we only included patients who had not evaluated an earlier prototype.

2A: Evaluate Intervention Flow and Content

Perceptions, needs, and preferences regarding the intervention flow and structure were evaluated with both patients with COPD (n=6) and HCPs (n=6) to identify usability requirements (Figure 1; phase 2A). Individual semistructured interviews were held using low-fidelity paper prototypes. The following topics were discussed: experience with mHealth and written action plans, the overall flow of the intervention, symptom-monitoring/action plan scenarios, and the added value of the intervention. Purposive sampling of participants was performed in primary, secondary, and tertiary care settings. In total, 6 patients and 6 HCPs were included based on the general rule of thumb that approximately 80% of all potential usability problems could be identified by including 5 to 10 end users [47]. The inclusion and exclusion criteria of the participants during usability testing are detailed in Textboxes 2 and 3, respectively. Data were thematically analyzed by two researchers independently [48]. Data analysis was supported by NVivo 10.0 software (2012; QSR International Pty Ltd.).

Textbox 2. Inclusion criteria of participants during usability testing.

Inclusion criteria for patients with a clinical diagnosis [2] of chronic obstructive pulmonary disease (COPD)

- Age >40 years
- Spirometry forced expiratory volume in 1 second/forced vital capacity ratio <70%
- ≥1 exacerbation in the last 12 months before entering the study (defined as a period of symptom deterioration in which the use of a course of corticosteroids and/or antibiotics was required, or hospitalization was necessary)
- Adequate communication skills
- Willing and able to comply with study procedures and give written informed consent
- Patients who are judged by their health care provider to have suitable hearing and vision

Inclusion criteria for health care providers

- Having a patient–health care provider relationship with patients with COPD
- Supporting patients with COPD in self-management
- At least one year of experience with COPD care

Textbox 3. Exclusion criteria of participants during usability testing.

Exclusion criteria for patients with a clinical diagnosis of chronic obstructive pulmonary disease

- Diagnosed with cognitive impairments
- Life expectancy ≤3 months
- Primary diagnosis of asthma, cardiac disease, or other major functionally limiting diseases

Exclusion criteria for health care providers

- Not applicable

2B: Evaluate Intervention Design Style

Next, the preferences of patients with COPD regarding intervention design style were explored by individual semistructured interviews (n=11; Figure 1; phase 2B) [33]. Low-fidelity paper prototypes were used to present variations in design style and tone of voice. Purposive sampling of participants was performed in a physiotherapy practice and a rehabilitation center according to the inclusion and exclusion criteria for patients in Textboxes 2 and 3. Data were analyzed by 2 researchers independently through summarizing the advantages and disadvantages of each design style and the overall preferences regarding design style. On the basis of the results of both alpha usability steps, the intervention design was finalized for further software development.

Phase 3: Iterative Software Development

The software of the mHealth intervention was developed during a 12-week period according to a scrum-based design method consisting of five development sprints (Figure 1; phase 3) [49]. During biweekly stakeholder meetings, the research team, designers, and software developers met in person to evaluate the current stage of development and to make decisions on the further development of the first version of the mHealth intervention (minimum viable product; MVP). The mHealth intervention was built in React Native (Massachusetts Institute of Technology licenses), a software structure that is easy to adapt and suitable for both iOS and Android. This saves time and money during the initial and future development of the intervention and fits within the agile development process of the intervention.

Phase 4: Field Usability Testing

In the fourth phase, field usability tests of the MVP (ie, tests with a high-fidelity prototype within the context in which the intervention will actually be used) were performed with patients with COPD (n=7) and HCPs (n=3) using cognitive task analysis [47,50]. This mixed methods study focused on three quality components: task success, user errors/problems, and satisfaction, based on Nielsen's heuristics and the International Organization for Standardization's usability standard 9241-11 [51]. Purposive sampling of participants was performed in primary, secondary, and tertiary care settings until data saturation was reached. In line with the procedure of phase 2, a minimum of 5 patients were included according to the inclusion and exclusion criteria of Textboxes 2 and 3. Participants were observed while performing tasks with the MVP and asked to *think aloud* to clarify their decision-making process and express experienced user problems and errors [51]. After the task analysis, the validated 10-item system usability scale (SUS) was filled out by patients to get a global view of usability [52]. Each item was scored on a 5-point Likert scale, and all items were converted to a total score (range 0-100, a score>70 is considered to be acceptable) [52,53]. Furthermore, semistructured interviews were conducted. On the basis of previous research and the technology acceptance model [54,55], the following topics were formulated: the first impression of the app, ease of use, satisfaction, perceived usefulness, applicability, attitude toward using the app, and the content of the app. The whole procedure with patients was video recorded without the faces of participants being visible. The procedure with HCPs was more pragmatic in nature because the MVP did not include a specific HCP interface. However, the relevant functions for HCPs could be tested within the MVP. Therefore, only 3 HCPs were included, and the procedure was only observed by 1 researcher who simultaneously made notes.

The performance of tasks by patients was observed by two researchers independently. An observation list was used to note task success, users' errors/problems, and participants' expressions for each task. The performance of tasks was scored as successful (1 point), partially successful (0.5 points), or unsuccessful (0 points) [56]. The observation lists were discussed by the researchers to reach a consensus on the performance of tasks and the identified problems and errors. The data from the think-aloud method were used to derive a better understanding of task performance. A severity score ranging from 0 (no usability problem) to 4 (usability catastrophe) was given to each problem based on the impact and frequency of the problem [57]. Data from the semistructured interviews were analyzed by 2 researchers independently using thematic analysis [48]. The data analysis of HCPs observations was performed by only 1 researcher, and the semistructured interviews were only summarized.

The usability studies were approved by the Medical Ethics Research Committee of the University Medical Center Utrecht (17–887), and all participants gave written informed consent.

Business Modeling

Business modeling, based on the principles of *the lean startup* methodology [58], was performed parallel to phase 1F until phase 4 to ensure valorization and sustainable implementation of the mHealth intervention in its intended care practice (Figure 1) [59]. Business modeling included contextual inquiry and continuous investigation of relevant stakeholder needs (patients with COPD, HCPs, policy makers, and health care insurers) to better understand what should be accomplished with our mHealth intervention and to obtain value drivers to underpin choices in what to design [59]. The needs of patients with COPD and HCPs were investigated in phases 1C, 1D, and 1E, and individual conversations with policy makers and health care insurers were held to identify their perspectives toward the mHealth intervention. Furthermore, the best innovation and distribution routes and market opportunities were explored in conversations with stakeholders to investigate their interests and financial incentives to support self-management with mHealth. Competition analysis was performed to explore the value of our intervention with respect to existing mHealth technologies. Finally, conversations with vendors in the field were held to explore business opportunities.

Results

Phase 1: Background Analysis and Design Conceptualization

1A: Identify the Evidence Base

A total of four relevant systematic reviews on exacerbation-related self-management interventions and two systematic reviews specifically focusing on mHealth interventions to improve exacerbation-related outcomes were identified. Self-management interventions, including exacerbation action plans along with ongoing self-management support, were associated with positive outcomes on quality of life, hospital admissions, and health care use [11,12,60]. A review of self-management interventions delivered immediately following an acute exacerbation showed no significant effect on quality of life nor hospital admissions [61]. All reviews showed large heterogeneity in interventions making it hard to draw conclusions on effective components of these interventions. Furthermore, mHealth interventions facilitating, supporting, and sustaining self-management among people with COPD significantly improved quality of life, and levels of activity [19]. Smartphone interventions in patients with COPD with exacerbations, without a specific focus on self-management, were found to be useful in reducing the number of patients having a COPD exacerbation [20]. These results should also be interpreted with caution because of the heterogeneity among studies.

On the basis of these findings, it seemed promising to use mHealth strategies that specifically aim at enhancing self-management behavior. It was considered important that the mHealth intervention includes a COPD action plan along with ongoing self-management support, which confirmed our guiding principle to include an action plan. Furthermore, a conceptual definition of a COPD self-management intervention was published in 2016 [16]. Given the need for consensus on what defines a COPD self-management intervention, this definition was added to the guiding principles (Textbox 1).

1B: Identify Key Behaviors and Concept Mapping

A conceptual model picturing the event of an exacerbation was developed by distinguishing five phases before, during, and after an index event. Specific aims regarding the reduction of exacerbation impact were formulated for each phase of the conceptual model. The conceptual model is published elsewhere [15]. On the basis of the knowledge generated from the literature, an initial set of 27 relevant self-management behaviors aiming to reduce exacerbation impact was identified and assigned to the relevant phases of the conceptual model. This initial set of self-management behaviors was introduced to experts in the first round of the Delphi study (Figure 1; phase 1C) to reach a consensus on the most relevant behaviors [15].

1C: Problem Analysis

A Delphi panel of 19 international experts reached a consensus on 17 self-management behaviors that can be targeted and influenced before, during, and after an exacerbation (Figure 1; phase 1C). This set of behaviors has the potential to maximally reduce the impact of exacerbations. The self-management behaviors were related to the following broader categories: adherence to pharmacotherapy, influenza vaccination, physical activity/exercise, avoiding stimuli, smoking cessation, early detection of symptom deterioration, medical treatment of exacerbations, managing stress and anxiety, and awareness of recurrent exacerbations [15]. The 17 self-management behaviors were considered as potential target behaviors for the mHealth intervention. Our grounded theory study (Figure 1; phase 1C) has resulted in a conceptual model explaining factors that influence exacerbation-related self-management from the patients' perspective. The conceptual model is published elsewhere [30]. The conceptual model shows that exacerbation-related self-management is influenced by five generic factors: *acceptance of COPD*, *perceived severity of symptoms*, *knowledge of exacerbations*, *former experiences with exacerbations*, and *social support*. Furthermore, *heterogeneity of exacerbations* and *habituation to symptoms* were identified as specific factors influencing the capability to recognize an exacerbation. Performance of self-management actions was specifically influenced by *perceived influence on exacerbation course*, *feelings of fear*, *self-empowerment*, *trust in health care provider*, *patient beliefs*, and *ambivalence toward treatment* [30]. These factors were included as moderating and mediating factors in the working mechanism of the intervention (see also 1F: Intervention Design section).

1D: Needs Analysis

Our needs analysis (Figure 1; phase 1D) resulted in an overview of potential benefits and barriers regarding the use of mHealth to support self-management and early ideas on the content of the intervention [31]. Both patients and HCPs emphasized the need for a multicomponent and tailored mHealth intervention that focuses on improving patient self-management skills by determining health status and providing adequate information, decision support, and feedback on self-management behavior in an advisory manner. Important findings were that patients and HCPs emphasized that an mHealth intervention should never replace patients' own feelings nor undermine their own decisions. The intervention should be complementary to regular (personal) contact with HCPs and should facilitate adequate self-management support by HCPs. Discussing self-management skills with HCPs in personal consultations was believed to be essential to improve these skills. Both patients and HCPs expressed doubts regarding (real-time) the monitoring of symptoms by HCPs because of safety reasons and time constraints, although early detection of exacerbations was considered to be an important benefit. Moreover, the intervention should be attractive, straightforward, rewarding, and safe. Finally, patients emphasized that using mHealth should be their own choice and should never be enforced. On the basis of these findings, the design requirements for the intervention were formulated. Further results of the focus group interviews are published elsewhere [31].

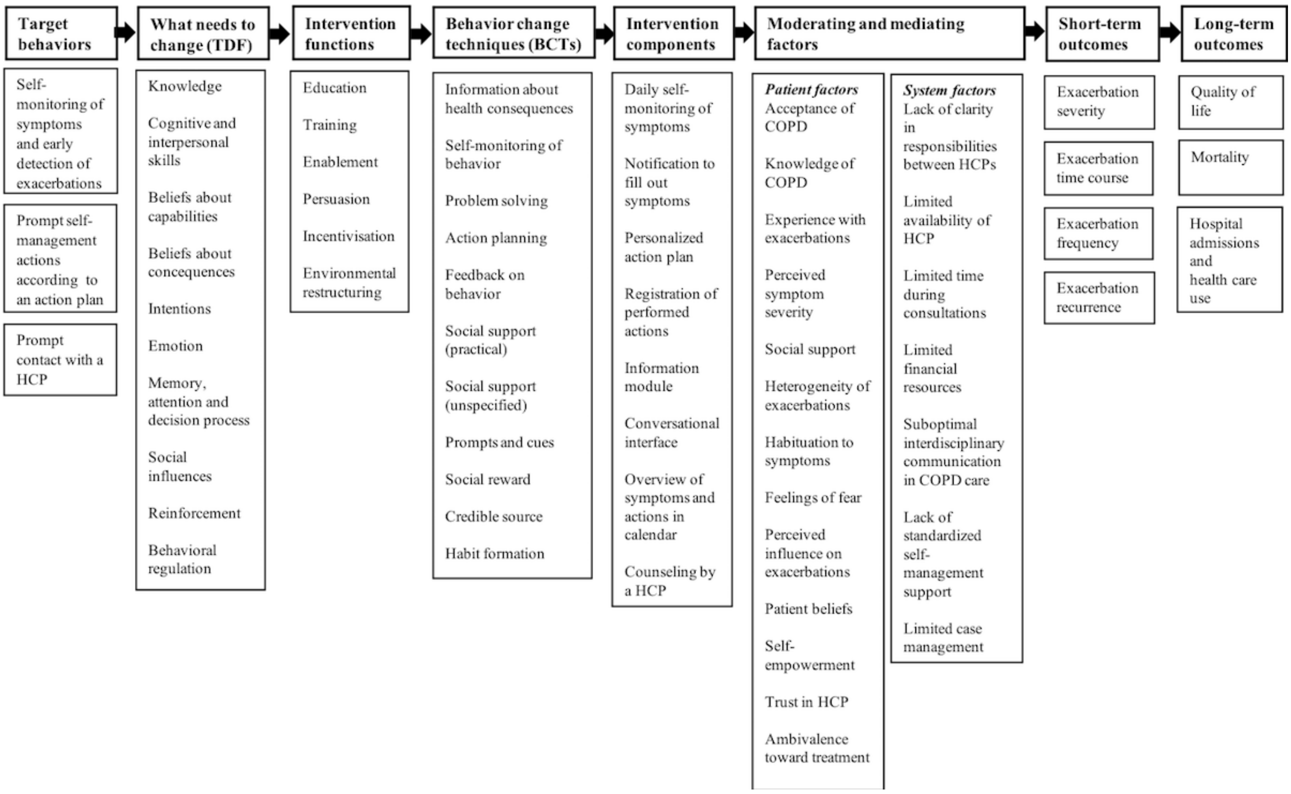
1E: Current Practice Analysis

On the basis of three Dutch health care standards focusing on COPD [62-64] and 10 interviews with HCPs, insight into current exacerbation-related self-management support was provided. Two pulmonologists, 2 nurse specialists, 2 pulmonary nurses, 2 general practitioners, and 1 primary care nurse (4 males/6 females, work experience range 3-20 years) were interviewed. An important finding was the lack of standardized self-management support and limited use of evidence-based interventions by HCPs. There was a large variation in providing information about exacerbation-related self-management with regard to timing, topics discussed, and mode of delivery. Only a few HCPs used a COPD action plan and prescribed self-treatment with prednisolone and/or antibiotics to stimulate self-management. More than half of the HCPs (n=6) expressed that patients had no specific case manager. Providing self-management support was mostly perceived as a shared responsibility between HCPs, although individual responsibilities of the HCPs involved were unclear. Most HCPs felt that there is large room for improvement in self-management support by HCPs. Barriers in providing self-management support were the lack of standardized self-management support and clarity in responsibilities between HCPs, limited availability of HCPs, limited case management, limited time during consultations, limited financial resources, and suboptimal interdisciplinary communication in COPD care. The findings from the current practice analysis were included as moderating and mediating factors in the working mechanism of the intervention (see also 1F: Intervention Design section).

1F: Intervention Design

During the intervention design phase, the research team decided to initially focus on the three target behaviors: (1) self-monitoring of symptoms and early detection of an exacerbation, (2) taking prompt individualized self-management actions, and (3) prompt contact with an HCP. On the basis of insights from phase 1C, these behaviors were expected to contribute most to the reduction of exacerbation impact, had the potential for large improvement, and were considered most feasible to influence. The choice for these three behaviors was made, given the importance of aggregating the target behaviors that fit together and are considered to have the largest impact on exacerbations [65]. On the basis of behavioral analysis of these behaviors, potential intervention functions and BCTs were selected for the intervention [41]. The behavioral analysis of the target behaviors, including the final intervention functions and BCTs is described in detail in Multimedia Appendix 1. Figure 2 shows the logic model of the intervention that synthesizes all the evidence gained in the previous phases (phase 1A until 1E), including the selection of final intervention functions and BCTs.

Figure 2. Logic model of a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease. BCTs: behavior change techniques; COPD: chronic obstructive pulmonary disease; HCP: health care provider; TDF: theoretical domains framework.



On the basis of the results of all previous phases, design requirements for the mHealth intervention were formulated (see Textbox 4). At this stage, the research team and design agency decided to develop a mobile app to enhance exacerbation-related self-management in patients with COPD. On the basis of design requirements, a concept of the flow and content of the app and various design styles were developed using low-fidelity paper prototypes.

Textbox 4. Design requirements for the mobile health intervention.

The mobile health (mHealth) intervention should:

- at least focus on self-monitoring of symptoms and early detection of exacerbations and taking prompt self-management actions including prompt contact with a health care provider (HCP);
- support patients in developing self-management skills over time (*learning by doing*) and changing behaviors;
- focus on (aggregated) self-management behaviors before, during, and after an exacerbation;
- include a chronic obstructive pulmonary disease (COPD) action plan with an educational component along with ongoing support;
- be comprehensive/multicomponent and tailored to individual patients;
- provide adequate information, decision support, and feedback on self-management behavior (in an advisory manner);
- take into account the factors influencing exacerbation-related self-management;
- fit with current COPD care and be accessible for HCPs;
- facilitate adequate self-management support by HCPs;
- be complementary to regular (personal) contact with HCPs;
- be attractive, straightforward, rewarding, and safe;
- never replace patients' own feelings nor undermine their own decisions; and
- not be enforced to patients. Using mHealth should be the patient's own choice.

Furthermore, the symptom-monitoring module was developed during this stage. Content validity of the symptom-monitoring module was determined after two expert rounds. In total, these eight symptoms were rated as relevant (I-CVI>0.78) and clear (≥75% of the expert panel): Dyspnea, wheezing, nighttime symptoms, coughing, sputum volume, sputum purulence, sputum color, and fatigue. The relevance of the final symptom-monitoring module, determined by three questions, was considered to be high with S-CVIs of 0.93 or greater. More detailed results and the final symptom-monitoring module are shown in Multimedia Appendix 2.

Phase 2: Alpha Usability Testing

2A: Evaluate Intervention Flow and Content

Evaluation of the flow and content of the intervention with both patients with COPD and HCPs resulted in overarching themes related to the intervention flow and an overview of usability requirements. The baseline characteristics of the participants are shown in Multimedia Appendix 3. The patients recruited from tertiary care all had a written action plan, whereas the other patients only had verbal agreements with their HCPs.

The Intervention Flow and Content

Overall, all patients and HCPs were positive about the intervention flow that consisted of four steps: (1) personalization of an action plan, (2) intensive monitoring of symptoms, (3) adjusting initial action plan based on monitoring period, and (4) regular use (filling out symptoms on a regular basis and receiving support on individualized actions). However, the patients who believed that they were well aware of their symptoms did not directly perceive that they could benefit from the intensive monitoring period. Overall, patients preferred personalization of the duration of intensive monitoring and the timing of notifications. HCPs felt that they should have autonomy in determining how, and at which moment, the action plan should be reviewed and adjusted. On the basis of specific mockups used to explore preferences regarding symptom registration and determining symptoms status, the most intuitive and straightforward scenarios were identified. For example, mockup 1 was considered to be the best solution to determine symptom status by all patients (Figure 3).

Figure 3. Low-fidelity paper prototypes of 3 scenarios to determine symptoms status.

Mockup 1 Mockup 2 Mockup 3

The Added Value of the Intervention

Patients expressed that the app could create awareness into their own situation and could support early detection of symptom deterioration and taking prompt actions. The adjustability and accessibility of the app were perceived as benefits compared with using a written action plan. Furthermore, both patients and

HCPs were positive about the overview of registered symptoms and undertaken actions as a tool to start the dialog about patients' self-management behavior. Moreover, personalization and tailoring of the app were considered to be an important benefit:

> *Finally an app that is not for COPD but for me personally.* [Patient 2]

Needs With Regard to the Intervention

Patients expressed a clear need for an accessible and reliable app that provides insight into their own situation and eliminates their doubts by including reflective questions. Patients stressed the importance of an app that is straightforward, for example, by providing simple and effortless instructions in case of serious dyspnea that causes panic, such as:

> *breathe slowly or call the doctor: It has to be simple, because energy is air.* [Patient 3]

They would like to use the app to inform relatives about their situation. Both patients and HCPs emphasized that the app should stimulate prompt contact with an HCP:

> *Patients experience feelings of fear you know, like: when I am raising an alarm, I might have to take prednisolone or I might be admitted to the hospital, so therefore I won't make the call...Will that be included in the app as well?* [HCP 1]

Furthermore, HCPs explained that the app should provide insight into patient symptoms over time and realize more proactive care instead of reactive care. Most HCPs felt that a separate HCP interface to personalize the app would increase the usability of the app in daily practice.

Usability Requirements

On the basis of the results of phase 2A, usability requirements were formulated for the software development phase (see Table 1).

Table 1. Usability requirements for the mobile health intervention.

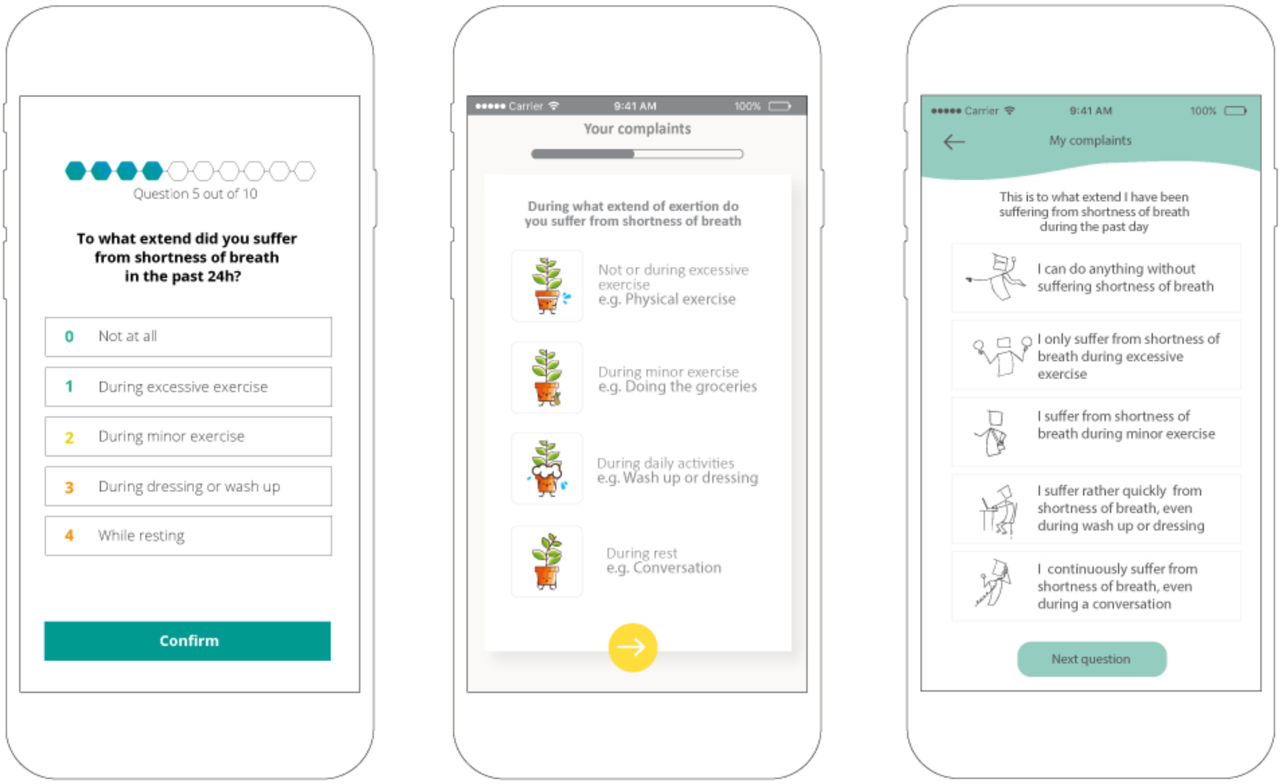
Topics of importance to users	Patients with chronic obstructive pulmonary disease (n=6) and HCPs[a] (n=6)
Content	• App should be reliable and accessible • Information should be straightforward and individualized • App should provide insight into symptoms over time • App should support prompt contact with an HCP • App should be a tool that can be used to get into dialog with HCPs about self-management behavior
Tailoring of the app	• Intensive monitoring period and timing of notifications should be personalized • HCPs should have autonomy in determining how, and at which moment in the care process, the action plan should be reviewed and adjusted
Interface	• The action plan should preferably be personalized in a separate HCP interface
Design style	• The design style should be restrained and clear without too much text.

[a] HCP: health care provider.

2B: Evaluate Intervention Design Style

In total, three potential design styles of the intervention were explored with patients with COPD (see Baseline characteristics in Multimedia Appendix 3. Mockups of the action plan, symptom registration, action registration, and the overview of symptoms and actions over time were used (see example in Figure 4). There was no consensus on a preferred design style. In general, patients preferred a restrained and clear design style without too much text. A few patients were positive about a more numerical design, whereas other patients found it hard to express their symptoms in numbers. There was a wide variety in preference regarding the tone of voice (distant vs personal tone of voice). Overall, patients were positive about using symbols. Most patients were negative about using an avatar in the app as it has no added value, and some patients considered an avatar to be childish. On the basis of these results, a restrained and clear design style without too much text was included as a usability requirement (see Table 1).

Figure 4. Low-fidelity paper prototypes of three symptom monitoring design styles.



Question 5 out of 10

To what extend did you suffer from shortness of breath in the past 24h?

0 Not at all
1 During excessive exercise
2 During minor exercise
3 During dressing or wash up
4 While resting

During what extend of exertion do you suffer from shortness of breath

Not or during excessive exercise e.g. Physical exercise
During minor exercise e.g. Doing the groceries
During daily activities e.g. Wash up or dressing
During rest e.g. Conversation

This is to what extend I have been suffering from shortness of breath during the past day

I can do anything without suffering shortness of breath
I only suffer from shortness of breath during excessive exercise
I suffer from shortness of breath during minor exercise
I suffer rather quickly from shortness of breath, even during wash up or dressing
I continuously suffer from shortness of breath, even during a conversation

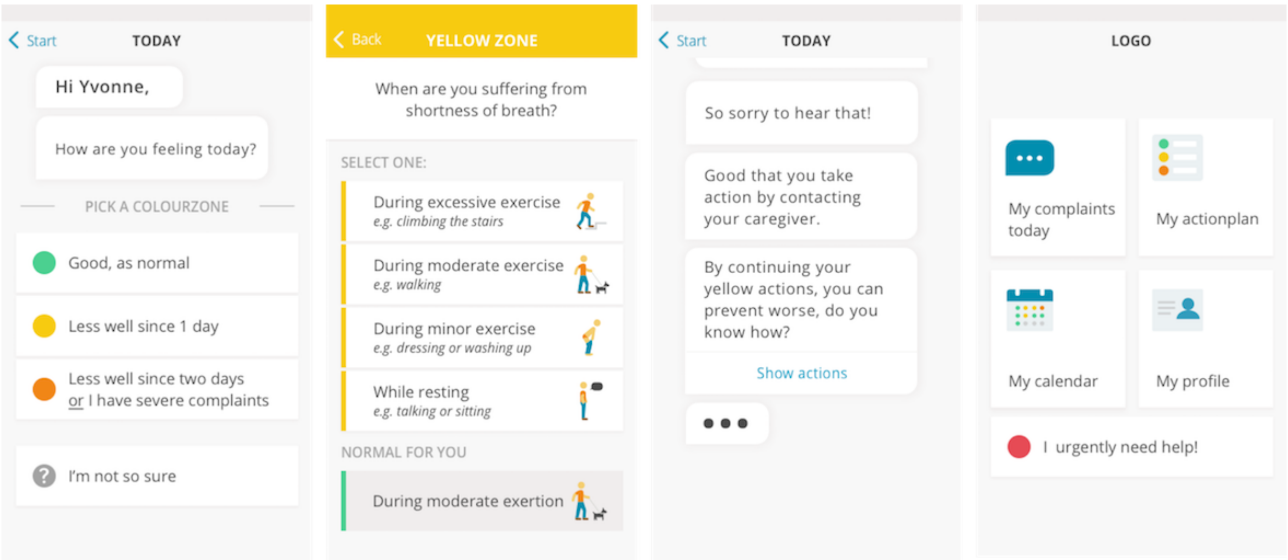
Phase 3: Iterative Software Development

Iterative software development resulted in a functional mobile app for patients (Copilot app) that can be used to (1) compose an action plan together with an HCP (based on a COPD action plan using color zones that is included in Dutch care standards)[66], (2) monitor symptoms and undertaken actions, (3) review symptoms and undertaken actions, and (4) read information about COPD and exacerbations. At this stage, the final integrated list of BCTs needed to ensure this MVP adhered to both the guiding principles (Textbox 1) and the design requirements (Textbox 4) was constructed by two researchers (YK and SH) using the BCW framework. In total, 6 intervention functions and 11 BCTs were selected for the MVP (see Figure 2 and Multimedia Appendix 1). During the iterative software development process, the research team made decisions to add steps to the flow of the app that were not thought of beforehand, such as including an onboarding program to register and personalize the patient's action plan. At the same time, owing to time and financial constraints, some steps were disregarded and moved to later versions of the intervention, such as including assistance in cases where patients are in doubt about contacting their HCP.

Phase 4: Field Usability Testing

All functionalities of the MVP were tested by patients with COPD, and all functionalities that belong to the HCP role were evaluated with HCPs (see baseline characteristics in Multimedia Appendix 3). Examples of the MVP are shown in Figure 5.

Figure 5. High-fidelity prototype of the Copilot app (minimum viable product).


Hi Yvonne,

How are you feeling today?

When are you suffering from shortness of breath?

So sorry to hear that!

Good that you take action by contacting your caregiver.

By continuing your yellow actions, you can prevent worse, do you know how?

Usability Assessment

According to patients with COPD, the general usability of the MVP was considered to be good based on the average rating of 90.7 (SD 6.7) on the SUS.

Task Success

Key tasks within the app were performed by all patients, and some tasks were performed only by a few patients because of time constraints (see Multimedia Appendix 4). Almost all patients were successful in consulting the app for HCP contact details and immediate help as well as to check self-management actions in the specific color zones. Furthermore, reviewing symptoms and undertaken actions was performed successfully overall, except for 2 participants who did not know where to find this overview. The information module was easily found by most of the patients. Most difficulties in the performance of tasks were observed during the symptom and action registration. Most patients were able to fill out their symptoms, but some patients overlooked the *save* button. One patient experienced difficulty with selecting the right color zone because of difficulty with scenario thinking. The support option for selecting a color zone (gray zone) appeared not to be intuitive, and navigation problems in the gray zone were observed. Overall, the HCPs were able to personalize the action plan and to evaluate the symptom-monitoring period. The performance of tasks by both patients and HCPs is further specified in Multimedia Appendix 4.

User Errors and Problems

On the basis of task analysis and observations during random navigation in the app, 23 user errors and problems were identified by patients with COPD. In total, seven problems were rated with the highest severity score of 4. These problems were related to saving registered symptoms, accidentally deleting symptoms and navigation in the gray zone. Moreover, five problems had a severity score of 3 and were related to the action plan overview, understanding of buttons on the home screen, and changing HCP contact details. The lowest severity scores (1 and 2) were assigned to 11 problems. Furthermore, 11 user errors and problems were observed during the use of the app by HCPs. The HCPs experienced problems with saving registered symptoms as well. Another severe problem was related to changing contact details (severity score 4). Two less severe problems were related to personalizing and changing the yellow zone of the action plan (severity score 3). The lowest severity scores (0-2) were assigned to seven problems. A more detailed overview of all user problems and errors is provided in Multimedia Appendix 4.

Patients' and Health Care Providers' Perceptions Toward Using the App

Overall, patients were positive about the app, as they found the app supportive with regard to monitoring and evaluating symptoms and taking prompt actions. On a scale from 1 to 10, patients' satisfaction rates ranged from 8 to 10 because of the ease of use and the interface being intuitive. The patients who frequently experienced exacerbations expressed an important need for the app. Two patients who rarely experienced exacerbations explained that the app will only have an added value if additional support would be provided during the stable phase (green zone). The interviews resulted in 13 themes that were categorized into feelings about the app, the added value of the app, the content of the app, and facilitators and barriers to use the app. A description of these themes and illustrative quotes is provided in Multimedia Appendix 4.

Field usability testing showed that a usable mHealth intervention has been developed. On the basis of these results, improvements for future versions of the app were revealed. The improvements were determined in collaboration with the designers and focus on the user errors and problems that were rated as most severe (severity ratings 3 and 4), problems that were observed in both patients and HCPs, or additional problems that were mentioned in the interviews. An overview of these improvements is shown in Multimedia Appendix 4.

Business Modeling

Business modeling resulted in a preliminary business plan that provided important design input during the development steps and will direct future development and implementation steps.

The business plan guided the direction to actively involve HCPs in providing the intervention to patients and the specific focus on developing self-management skills over time (learning by doing) to be distinctive from other Dutch mHealth solutions. These outcomes were added to the design requirements (Textbox 4). During business modeling, market volume, and segmentation, different innovation and distribution routes and revenue models were systematically evaluated. Given current positive developments with regard to funding of apps in the Netherlands, especially those that are evidence-driven, a distribution strategy will be chosen that includes health care insurers to ensure implementation and continued use in Dutch COPD care.

Discussion

Principal Findings

This paper provides insight into a systematic and thorough way of developing an evidence-driven and usable mHealth intervention (Copilot) for patients with COPD to enhance exacerbation-related self-management. Following an iterative UCD process, a mobile app consisting of a personalized action plan and symptom-monitoring module has been developed. The intervention was developed by following a thorough and well-underpinned process consisting of a background analysis and design conceptualization phase leading to final guiding principles, a logic model and design requirements, usability testing phases leading to usability requirements, and iterative software development of an MVP that adheres to these guiding principles and design and usability requirements. This unique approach of scientific engineering has resulted in an mHealth intervention that meets the needs and preferences of patients with COPD, is likely to be used by patients with COPD, and has a high potential to be effective in reducing exacerbation impact. Involving patients with COPD, HCPs, COPD experts, and experts from design and behavioral science throughout the development process increased the likelihood that the mHealth intervention can be successfully implemented into Dutch COPD care.

Copilot requires an active case manager role, as previous studies have shown the need for ongoing case manager support alongside the use of an action plan to achieve effective and safe self-management [11,67]. The mHealth intervention was developed as a tool to enhance patient self-management skills that is complementary to personal interaction with an HCP. This is in line with recent research underlining that a good patient-HCP relationship is important for patients to engage and take responsibility for their own health care [68,69]. The specific focus on developing self-management skills over time is distinctive from other mHealth initiatives, as research in the past decade has focused increasingly on telemonitoring strategies to decrease the impact of exacerbations [21,70,71]. The impact of telemonitoring in the COPD population is, however, still equivocal because of trial designs, unstandardized interventions, and limited follow-up [21,70]. With telemonitoring, the decision-making process is profession based. The working mechanism of Copilot focuses on enhancing patients' self-management skills over time. Therefore, no telemonitoring strategies were included in our mHealth intervention.

Although our Delphi study has shown the need for a comprehensive strategy to improve the full spectrum of exacerbation-related self-management behavior [15], the first version of our mHealth intervention focuses specifically on self-monitoring of symptoms and taking prompt individualized self-management actions. A *less is more* approach consisting of only a few strong target behaviors that fit together was considered to be imperative in creating impact [41,65]. When changing these target behaviors is proven effective, we could build upon these behaviors incrementally [41]. It is important to note that not all relevant self-management behaviors have to be addressed through mHealth, as HCPs should continue to have an essential role in providing self-management support as well [31].

Strengths and Limitations

A major strength was the systematic and thorough way of developing Copilot according to a UCD that was based on existing development models and diminished the chance of missing important steps [33,34]. We have systematically investigated and incorporated the views of end users, continuously evaluated prototypes, and involved persuasive design techniques to match user profiles and motivate patients to engage in self-management, which is in line with the *person-based approach* and the holistic framework for the development of electronic health (eHealth) technologies [32,72]. Furthermore, we used guiding principles to easily recall the principal and distinctive features of the intervention during the extensive, iterative intervention development process [32]. Another important strength was the detailed analysis of behaviors of patients with COPD using the BCW method and selection of BCTs to underpin the pathway toward behavior change [41]. Using the BCW method along with a UCD is comparable with the methodology used by Curtis et al [22] to develop a theory-driven and user-centered healthy eating app. Their work also focused on a thorough analysis of target behaviors, selection of BCTs, and exploration of user preferences to underpin the design of the app with relevant theory and evidence and ensure engagement among the target population. However, Curtis et al [22] performed no specific activities with regard to valorization and implementation of their app during their development process [22]. To make both the design and the implementation of our app value driven, we performed valorization activities throughout the development process of our app [59,72]. Business modeling helped us to identify critical success factors that will influence the sustainability and effectiveness of the app, which is often overlooked during the development process of eHealth and mHealth technologies [59].

From a health care and behavior change perspective, we chose to use the MRC framework for the development of complex interventions as a basis for our UCD, instead of a more general software development approach. The four iterative phases were inspired by the user-centered methodology used by Johnston et al [33] for the development of a Web-based interface for patients with COPD. The use of a more general software development approach as a basis for the development process might have provided more specific guidance to the software development and usability phases beforehand. However, such approaches pay less attention to the activities needed to design a theory- and evidence-driven intervention, which was an important focus in our design process.

Furthermore, it should be noted that the extensive and thorough development process increased the likelihood of developing an effective intervention, although it is questionable if this process is feasible in daily practice. The whole intervention development process took place over a 4-year period, which is quite time consuming and could increase the risk of a misfit with current market developments or that technology has moved on by the time of implementation [22,73,74]. The time-consuming development was partly because of the inclusion of all the development phases but was also related to developing an mHealth intervention from a scientific environment. Developing an intervention from science involves completing an empirical cycle at each development phase and often includes an extensive review of a study protocol by a medical ethics research committee. Pursuing the rules of science during the development process of an mHealth intervention has slowed down the process at certain points in time. Furthermore, development from a scientific environment generally means less focus on business modeling and entrepreneurship, which could delay the process of bringing the app to the market. Finally, a limitation in this study was the restricted budget available for the creative design and development of the mHealth intervention, which required us to make a selection in the development of intervention components.

Implications for Practice and Future Research

The findings of this study are important for both patients with COPD and HCPs supporting patients with COPD in self-management as well as for researchers and designers focusing on the development of mHealth interventions. For patients with COPD, an evidence-driven and usable mobile app has been developed to assist them in developing exacerbation-related self-management skills. It needs to be emphasized that, at least for the coming years, not all patients with COPD will be eligible for the mHealth intervention, especially for those with a more negative attitude toward mHealth and low digital literacy [19,31]. However, a positive change in attitudes toward mHealth and digital skills can be expected in the future, given the current trends in internet access and smartphone use [18]. For HCPs, the mHealth intervention can be used to provide more evidence-based, structured, and tailored self-management support. The mHealth intervention can be embedded in primary, secondary, and tertiary care settings, which could contribute to improving integrated care. To increase the likelihood of successful implementation in Dutch COPD care, the intervention can easily be adapted to a specific setting and context. Hereby, the intervention would be available for a wide range of health care settings in which patients with COPD are currently treated. For future practice, it is important that more intervention components will be added to the mHealth intervention to optimally address the selected target behaviors, such as adding self-treatment with prednisolone and/or antibiotics and providing assistance in case patients are in doubt about if they should contact their HCP. Furthermore, a separate dashboard for HCPs should be developed to be able to individualize the mHealth intervention and to review registered symptoms and actions during consultations without having to use the patient's own device. An essential step would then be to establish cooperation with external vendors in the field and health care insurers to ensure implementation in COPD care. In the next phase, it is important that the mHealth intervention takes into account patient comorbidities to make the intervention available for a wider population and to ensure patient safety [12]. Future steps should focus on adding target behaviors that are relevant before, during, and after an exacerbation to maximize the reduction of exacerbation impact.

For researchers and designers, the UCD in this study can be used as guidance for the development of mHealth interventions that meet end user needs and preferences, have high potential to be effective, and are likely to be used by the target population. Essential in the development is that interventions are grounded in theory and evidence and that user needs and preferences are thoroughly investigated. Moreover, valorization and implementation activities should be regarded as continuous activities throughout the development process to ensure sustainable use in its intended practice. This extensive reporting of the intervention development process enhances the reproducibility of the intervention and contributes to more transparency in the development of complex interventions in health care, which is needed to strengthen the internal and external validity of interventions and to add value to health care research [34]. All in all, it is helpful to have multiple examples and variants on how to develop evidence- and theory-driven mHealth interventions. It should be considered if the thoroughness of this UCD is needed for all mHealth interventions that will be developed in the future. Depending on the topic, decisions should be made about which phases and steps are relevant to the topic and should be included in the development process. In addition, taking time aspects into consideration, it should be questioned how thoroughly an individual step should be executed. The need for efficiency in the development of mHealth interventions is currently a highly discussed topic [75]. Our work contributes to this discussion by mapping out a state-of-the-art design and development process and showing how time consuming this is.

Future research should focus on evaluating the feasibility of the mHealth intervention in the daily practice of HCPs, as they have a key role in personalizing the mHealth intervention before patient use. In a second phase, the feasibility of the mHealth intervention should be evaluated with patients with COPD to investigate the delivery and acceptability of the intervention, compliance with the intervention, and recruitment and retention of patients. In the next phase, the effect of the mHealth intervention on the relevant patient outcomes and health care use should be evaluated. Recent studies on mHealth interventions in patients with COPD suggest the use of randomized controlled trials (RCTs) with adequately powered sample sizes and a 1-year follow-up period to be sufficient to comment on behavioral change and impact of treatment [19,20]. However, this time-consuming design may not be ideal for rapidly evolving mHealth technologies [73,76]. Using an RCT implies two or more years of research in which this mHealth intervention with high potential for effectiveness, and no expected harm will not be available for patients with COPD.

Furthermore, an RCT only enables identifying if this complex mHealth intervention as a whole work and the cost-effectiveness of it, without identifying which intervention components work in whom. Alternatively, more rapid study designs such as n-of-1 trials or observational designs could be used to understand the working mechanism of the intervention and simultaneously focus on bringing the mHealth intervention to the market as soon as possible [73,75-77]. Within these designs, it is important to evaluate self-management skills and behavior change as outcomes, and the way this is assessed should be clearly reported [19,77,78].

Conclusions

This paper described in detail the full UCD and development process of an evidence-driven and usable mHealth intervention to enhance exacerbation-related self-management in patients with COPD. By following a UCD process, an mHealth intervention was developed that meets the needs and preferences of patients with COPD, is likely to be used by patients with COPD, and has a high potential to be effective in reducing exacerbation impact. This extensive reporting of the intervention development process contributes to more transparency in the development of complex interventions in health care. The UCD process in this study can be used by researchers and designers as guidance for the development of mHealth interventions. However, taking time aspects into consideration, decisions have to be made about the thoroughness of executing individual phases.

Acknowledgments

The authors thank all patients and HCPs who participated throughout the development process for their time and sharing perceptions toward the mHealth intervention. The authors acknowledge all health care centers for recruiting patients. The authors thank Beatrijs van Hoof and Jochem Wilson from Panton BV for translating the input for the mHealth intervention into a creative design and thank developers from Inspire Innovation BV for developing the software of Copilot. Furthermore, the authors thank Annabeth van Stroe, RN, MSc, for her role in evaluating the intervention flow and content and also thank the students from the Bachelor of Nursing course of the University of Applied Sciences Utrecht for their role in examining current practice. During the development process, researchers JT and YK joined an innovation contest (Ureka Mega Challenge) at the University Medical Center of Utrecht. A grant of €50,000 (US $55,645) was received from this contest, which enabled the development of the first version of the Copilot app.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Behavioral analysis of target behaviors and final intervention functions and behavior change techniques.
[DOCX File , 46 KB - jmir_v22i6e15449_app1.docx]

Multimedia Appendix 2
Content validity of the symptom monitoring module.
[DOCX File , 35 KB - jmir_v22i6e15449_app2.docx]

Multimedia Appendix 3
Demographic characteristics of participants during usability testing phases (phases 2 and 4).
[DOCX File , 18 KB - jmir_v22i6e15449_app3.docx]

Multimedia Appendix 4
Results field usability testing.
[DOCX File , 33 KB - jmir_v22i6e15449_app4.docx]

References

1. World Health Organization. The Global Burden of Disease: Update 2004. Geneva: World Health Organization; 2008.
2. Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2018. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: 2018 Report URL: https://goldcopd.org/wp-content/uploads/2017/11/GOLD-2018-v6.0-FINAL-revised-20-Nov_WMS.pdf [accessed 2018-06-12]
3. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002 Oct;57(10):847-852 [FREE Full text] [doi: 10.1136/thorax.57.10.847] [Medline: 12324669]

4. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1998 May;157(5 Pt 1):1418-1422. [doi: 10.1164/ajrccm.157.5.9709032] [Medline: 9603117]
5. Miravitlles M, Ferrer M, Pont A, Zalacain R, Alvarez-Sala JL, Masa F, IMPAC Study Group. Effect of exacerbations on quality of life in patients with chronic obstructive pulmonary disease: a 2 year follow up study. Thorax 2004 May;59(5):387-395 [FREE Full text] [doi: 10.1136/thx.2003.008730] [Medline: 15115864]
6. Toy EL, Gallagher KF, Stanley EL, Swensen AR, Duh MS. The economic impact of exacerbations of chronic obstructive pulmonary disease and exacerbation definition: a review. COPD 2010 Jun;7(3):214-228. [doi: 10.3109/15412555.2010.481697] [Medline: 20486821]
7. Soler-Cataluña JJ, Martínez-García MA, Sánchez PR, Salcedo E, Navarro M, Ochando R. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. Thorax 2005 Nov;60(11):925-931 [FREE Full text] [doi: 10.1136/thx.2005.040527] [Medline: 16055622]
8. Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. Chest 2000 May;117(5 Suppl 2):398S-401S. [doi: 10.1378/chest.117.5_suppl_2.398s] [Medline: 10843984]
9. Bourbeau J, van der Palen J. Promoting effective self-management programmes to improve COPD. Eur Respir J 2009 Mar;33(3):461-463 [FREE Full text] [doi: 10.1183/09031936.00001309] [Medline: 19251792]
10. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
11. Howcroft M, Walters EH, Wood-Baker R, Walters JA. Action plans with brief patient education for exacerbations in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2016 Dec 19;12:CD005074 [FREE Full text] [doi: 10.1002/14651858.CD005074.pub4] [Medline: 27990628]
12. Lenferink A, Brusse-Keizer M, van der Valk PD, Frith PA, Zwerink M, Monninkhof EM, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Aug 4;8:CD011682 [FREE Full text] [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]
13. Jonkman NH, Westland H, Trappenburg JC, Groenwold RH, Bischoff EW, Bourbeau J, et al. Characteristics of effective self-management interventions in patients with COPD: individual patient data meta-analysis. Eur Respir J 2016 Jul;48(1):55-68 [FREE Full text] [doi: 10.1183/13993003.01860-2015] [Medline: 27126694]
14. Trappenburg J, Jonkman N, Jaarsma T, van Os-Medendorp H, Kort H, de Wit N, et al. Self-management: one size does not fit all. Patient Educ Couns 2013 Jul;92(1):134-137. [doi: 10.1016/j.pec.2013.02.009] [Medline: 23499381]
15. Korpershoek YJ, Slot JC, Effing TW, Schuurmans MJ, Trappenburg JC. Self-management behaviors to reduce exacerbation impact in COPD patients: a Delphi study. Int J Chron Obstruct Pulmon Dis 2017;12:2735-2746 [FREE Full text] [doi: 10.2147/COPD.S138867] [Medline: 28979116]
16. Effing TW, Vercoulen JH, Bourbeau J, Trappenburg J, Lenferink A, Cafarella P, et al. Definition of a COPD self-management intervention: International Expert Group consensus. Eur Respir J 2016 Jul;48(1):46-54 [FREE Full text] [doi: 10.1183/13993003.00025-2016] [Medline: 27076595]
17. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
18. World Health Organisation. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organisation; 2011.
19. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 May 23;5:CD011425 [FREE Full text] [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
20. Alwashmi M, Hawboldt J, Davis E, Marra C, Gamble J, Ashour WA. The effect of smartphone interventions on patients with chronic obstructive pulmonary disease exacerbations: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 Sep 1;4(3):e105 [FREE Full text] [doi: 10.2196/mhealth.5921] [Medline: 27589898]
21. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease: Cochrane Review and meta-analysis. Br J Gen Pract 2012 Nov;62(604):e739-e749 [FREE Full text] [doi: 10.3399/bjgp12X658269] [Medline: 23211177]
22. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR Mhealth Uhealth 2015 Jun 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
23. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
24. Hagger MS, Weed M. DEBATE: Do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019 Apr 25;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
25. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical

mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512 [FREE Full text] [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]

26. Sheeran P, Klein WM, Rothman AJ. Health behavior change: moving from observation to intervention. Annu Rev Psychol 2017 Jan 3;68:573-600. [doi: 10.1146/annurev-psych-010416-044007] [Medline: 27618942]
27. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
28. Conn VS, Hafdahl AR, Cooper PS, Brown LM, Lusk SL. Meta-analysis of workplace physical activity interventions. Am J Prev Med 2009 Oct;37(4):330-339 [FREE Full text] [doi: 10.1016/j.amepre.2009.06.008] [Medline: 19765506]
29. Michie S, Brown J, Geraghty AW, Miller S, Yardley L, Gardner B, et al. Development of StopAdvisor: A theory-based interactive internet-based smoking cessation intervention. Transl Behav Med 2012 Sep;2(3):263-275 [FREE Full text] [doi: 10.1007/s13142-012-0135-6] [Medline: 24073123]
30. Korpershoek Y, Vervoort S, Nijssen L, Trappenburg J, Schuurmans MJ. Factors influencing exacerbation-related self-management in patients with COPD: a qualitative study. Int J Chron Obstruct Pulmon Dis 2016;11:2977-2990 [FREE Full text] [doi: 10.2147/COPD.S116196] [Medline: 27932877]
31. Korpershoek YJ, Vervoort SC, Trappenburg JC, Schuurmans MJ. Perceptions of patients with chronic obstructive pulmonary disease and their health care providers towards using mHealth for self-management of exacerbations: a qualitative study. BMC Health Serv Res 2018 Oct 4;18(1):757 [FREE Full text] [doi: 10.1186/s12913-018-3545-4] [Medline: 30286761]
32. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015;1:37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]
33. Johnston SK, Nguyen HQ, Wolpin S. Designing and testing a web-based interface for self-monitoring of exercise and symptoms for older adults with chronic obstructive pulmonary disease. Comput Inform Nurs 2009;27(3):166-174 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c1d] [Medline: 19411946]
34. Bleijenberg N, Ginkel JM, Trappenburg JC, Ettema RG, Sino CG, Heim N, et al. Increasing value and reducing waste by optimizing the development of complex interventions: Enriching the development phase of the Medical Research Council (MRC) Framework. Int J Nurs Stud 2018 Mar;79:86-93 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.12.001] [Medline: 29220738]
35. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
36. Strauss AL, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Third Edition. Thousand Oaks, CA: Sage Publications; 2007.
37. van Meijel B, Gamel C, van Swieten-Duijfjes B, Grypdonck MH. The development of evidence-based nursing interventions: methodological considerations. J Adv Nurs 2004 Oct;48(1):84-92. [doi: 10.1111/j.1365-2648.2004.03171.x] [Medline: 15347414]
38. Shah SG, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
39. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research. Fourth Edition. Los Angeles: Sage; 2008.
40. Boeije H. Analysis in Qualitative Research. London: Sage Publications; 2009.
41. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Great Britain: Silverback; 2014.
42. Baxter SK, Blank L, Woods HB, Payne N, Rimmer M, Goyder E. Using logic model methods in systematic review synthesis: describing complex pathways in referral management interventions. BMC Med Res Methodol 2014 May 10;14:62 [FREE Full text] [doi: 10.1186/1471-2288-14-62] [Medline: 24885751]
43. West R, Michie S. A Guide to Development and Evaluation of Digital Behavior Change Interventions in Health Care. London: Silverback Publishing; 2016.
44. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
45. Wiklund M, Kendler JA, Strochlic A. Usability Testing of Medical Devices. Second Edition. Boca Raton, FL: CRC press; 2011.
46. Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35(6):382-385. [Medline: 3640358]
47. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
48. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
49. van Amstel DP, Heemskerk M, Renes RJ, Hermsen S. The value of agile methods in designing for behavioural change: a case study. Des J 2017;20(sup1):S681-S690. [doi: 10.1080/14606925.2017.1353014]

50. Kushniruk AW, Patel VL, Cimino JJ. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. Proc AMIA Annu Fall Symp 1997:218-222 [FREE Full text] [Medline: 9357620]
51. Nielsen J. Usability Engineering. Boston, MA: Academic Press; 1993.
52. Brooke J. Usability. 1986. System Usability Scale (SUS) URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2018-09-02]
53. Science Direct. System Usability Scale 2019 URL: https://www.sciencedirect.com/topics/computer-science/system-usability-scale [accessed 2019-05-23]
54. Crane D, Garnett C, Brown J, West R, Michie S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies. Front Public Health 2017;5:39 [FREE Full text] [doi: 10.3389/fpubh.2017.00039] [Medline: 28421175]
55. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
56. Nielsen J. Nielsen Norman Group. Success Rate: The Simplest Usability Metric 2001 URL: https://www.nngroup.com/articles/success-rate-the-simplest-usability-metric/ [accessed 2017-11-28]
57. Nielsen J. Nielsen Norman Group. 1994. Severity Ratings for Usability Problems URL: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [accessed 2017-11-21]
58. Ries E. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation To Create Radically Successful Businesses. USA: Crown Publishing Group; 2011.
59. van Limburg M, van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011 Dec 28;13(4):e124 [FREE Full text] [doi: 10.2196/jmir.1674] [Medline: 22204896]
60. Zwerink M, Brusse-Keizer M, van der Valk PD, Zielhuis GA, Monninkhof EM, van der Palen J, et al. Self management for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2014 Mar 19(3):CD002990 [FREE Full text] [doi: 10.1002/14651858.CD002990.pub3] [Medline: 24665053]
61. Harrison SL, Janaudis-Ferreira T, Brooks D, Desveaux L, Goldstein RS. Self-management following an acute exacerbation of COPD: a systematic review. Chest 2015 Mar;147(3):646-661. [doi: 10.1378/chest.14-1658] [Medline: 25340578]
62. Long Alliantie Nederland (LAN). [National Care Path COPD Lung Attack with Hospitalization]. Netherlands: Long Alliantie Nederland; 2019.
63. Nederlands Huisartsen Genootschap. 2015. NHG standaard COPD URL: https://www.nhg.org/standaarden/volledig/nhg-standaard-copd [accessed 2017-06-01]
64. Long Alliantie Nederland (LAN). [COPD Care Standard]. Netherlands: Long Alliantie Nederland; 2016.
65. Richards DA. Complex interventions and the amalgamation of marginal gains: A way forward for understanding and researching essential nursing care? Int J Nurs Stud 2015 Jul;52(7):1143-1145. [doi: 10.1016/j.ijnurstu.2015.01.013] [Medline: 25746362]
66. Trappenburg JC, Monninkhof EM, Bourbeau J, Troosters T, Schrijvers AJ, Verheij TJ, et al. Effect of an action plan with ongoing support by a case manager on exacerbation-related outcome in patients with COPD: a multicentre randomised controlled trial. Thorax 2011 Nov;66(11):977-984. [doi: 10.1136/thoraxjnl-2011-200071] [Medline: 21785156]
67. Effing T. Action plans and case manager support may hasten recovery of symptoms following an acute exacerbation in patients with chronic obstructive pulmonary disease (COPD). J Physiother 2012;58(1):60 [FREE Full text] [doi: 10.1016/S1836-9553(12)70076-0] [Medline: 22341386]
68. Luhr K, Holmefur M, Theander K, Eldh AC. Patient participation during and after a self-management programme in primary healthcare - The experience of patients with chronic obstructive pulmonary disease or chronic heart failure. Patient Educ Couns 2018 Jun;101(6):1137-1142. [doi: 10.1016/j.pec.2017.12.020] [Medline: 29290329]
69. Andersen IC, Thomsen TG, Bruun P, Bødtger U, Hounsgaard L. Between hope and hopelessness: COPD patients' and their family members' experiences of interacting with healthcare providers - a qualitative longitudinal study. Scand J Caring Sci 2018 Sep;32(3):1197-1206. [doi: 10.1111/scs.12565] [Medline: 29446474]
70. Goldstein RS, O'Hoski S. Telemedicine in COPD: time to pause. Chest 2014 May;145(5):945-949. [doi: 10.1378/chest.13-1656] [Medline: 24798834]
71. McKinstry B. The use of remote monitoring technologies in managing chronic obstructive pulmonary disease. Q J Med 2013 Oct;106(10):883-885. [doi: 10.1093/qjmed/hct068] [Medline: 23564633]
72. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
73. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
74. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]

75. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
76. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
77. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
78. Effing TW, Bourbeau J, Vercoulen J, Apter AJ, Coultas D, Meek P, et al. Self-management programmes for COPD: moving forward. Chron Respir Dis 2012 Feb;9(1):27-35. [doi: 10.1177/1479972311433574] [Medline: 22308551]

Abbreviations

BCT: behavior change technique
BCW: behavior change wheel
COM-B: capability, opportunity, and motivation model of behavior
COPD: chronic obstructive pulmonary disease
eHealth: electronic health
HCP: health care provider
I-CVI: item-content validity index
MEDLINE: Medical Literature Analysis and Retrieval System Online
mHealth: mobile health
MRC: Medical Research Council
MVP: minimum viable product
RCT: randomized controlled trial
S-CVI: scale-content validity index
SUS: system usability scale
TDF: theoretical domains framework
UCD: user-centered design

Edited by G Eysenbach; submitted 11.07.19; peer-reviewed by C McCallum, J Rooksby; comments to author 21.10.19; revised version received 12.11.19; accepted 24.01.20; published 15.06.20.

Please cite as:
Korpershoek YJG, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JCA
User-Centered Design of a Mobile Health Intervention to Enhance Exacerbation-Related Self-Management in Patients With Chronic Obstructive Pulmonary Disease (Copilot): Mixed Methods Study
J Med Internet Res 2020;22(6):e15449
URL: https://www.jmir.org/2020/6/e15449
doi:10.2196/15449
PMID:32538793

Original Paper

Challenges and Benefits of an Internet-Based Intervention With a Peer Support Component for Older Adults With Depression: Qualitative Analysis of Textual Data

Annie T Chen[1], MSIS, PhD; Krystal Slattery[1], MS; Kathryn N Tomasino[2,3], PhD; Caryn Kseniya Rubanovich[2,4], MS; Leland R Bardsley[2], MS; David C Mohr[2], PhD

[1]Department of Biomedical Informatics and Medical Education, University of Washington School of Medicine, Seattle, WA, United States

[2]Department of Preventive Medicine, Center for Behavioral Intervention Technologies, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Division of Gastroenterology & Hepatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[4]San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, United States

Corresponding Author:
Annie T Chen, MSIS, PhD
Department of Biomedical Informatics and Medical Education
University of Washington School of Medicine
850 Republican Street, Box 358047
UW Medicine South Lake Union
Seattle, WA, 98109
United States
Phone: 1 2062219218
Email: atchen@uw.edu

Abstract

Background: Technological interventions provide many opportunities for improving the health and quality of life of older adults. However, interaction with new technologies can also cause frustration. Although these themes have been explored in extant research, much remains to be learned with regard to how the challenges of aging and technology use and the experiences of participating in a social and learning environment are interrelated.

Objective: This study aimed to perform a qualitative analysis of data collected from MoodTech, a pilot study of an internet-based intervention with a peer support component for older adults with symptoms of depression, to better understand the participants' experience of using technological interventions, including the challenges and benefits that they experienced over the course of these interventions.

Methods: We employed an inductive qualitative analysis method based on grounded theory methodology and interpretative phenomenological analysis to analyze participant textual data. These textual data were of 3 main types: (1) assignments in which participants challenged their negative thoughts, (2) status updates, and (3) comments in the peer support component of the intervention.

Results: We have presented the results through 3 main themes: (1) the challenges of aging as seen through the participants' comments, (2) the difficulties experienced by the participants in using MoodTech, and (3) the benefits they derived from participating.

Conclusions: This paper offers several contributions concerning study participants' experiences with internet-based cognitive behavioral therapy (iCBT) interventions with a peer support component and design considerations for developing complex technological interventions that support the challenges participants experience due to aging and cognitive difficulties. First, technical issues encountered by older adults within the context of the intervention can interact with and exacerbate the insecurities they experience in life, and it is important to consider how intervention components might be designed to mitigate these issues. Second, peer support can be employed as a mechanism to facilitate communication, support, and collaborative problem solving among participants in an intervention. The insights from this paper can inform the design of iCBT interventions for older adults.

(J Med Internet Res 2020;22(6):e17586) doi:10.2196/17586

KEYWORDS

aged; depression; internet; peer group; social support; qualitative research

Introduction

Background

Depression is a psychiatric disorder that affects a significant proportion of older adults [1,2] and can have various negative consequences, including decreased quality of life, increased disability, and increased risk of mortality [3]. In addition, depression can often go undetected or untreated [2,4]. The treatments for depression in older adults include antidepressants, behavioral treatments, cognitive behavioral therapy (CBT), cognitive bibliotherapy, problem-solving therapy, psychotherapy, and life review/reminiscence therapy [2,5,6]. In recent years, there has been interest in internet-based CBT (iCBT) for reducing depressive symptoms in older adults; however, the number, quality, and heterogeneity of extant research suggest the need for additional research to evaluate the efficacy of iCBT interventions [7].

There has also been recognition of the negative effects of social isolation among older adults, including poorer cognitive functioning, impaired sleep and daytime dysfunction, reductions in physical activity, impaired mental health, and the potential utility of technology for alleviating loneliness and social isolation [8,9]. Although studies of technological interventions for depression and loneliness among older adults are still limited, extant research suggests that these interventions can reduce the symptoms of anxiety and depression [10,11] and are helpful for the management of loneliness [8,12,13].

However, older adults may experience difficulties in learning new technologies, and it is important to provide support during the process of learning to use them [14]. Although digital technologies could be a vehicle for empowerment by facilitating greater engagement in hobbies, social support, and everyday tasks, they may also be a form of disempowerment if they are viewed with apprehension, and the inability to use them can lead to social isolation [15]. Moreover, the adoption of technology use can be mediated by cognitive abilities, computer self-efficacy, and computer anxiety [16].

Nevertheless, previous research has suggested that older adults perceive that the benefits of using these technologies may outweigh the costs [17] and that engagement with internet and digital technologies can lead to improved depression symptomatology and psychological well-being [18,19]. In recent years, there has been increased interest in the factors that influence engagement with social networking websites among older adults [20]. For example, some older adults use Facebook to connect with family, and previous research has reported that older adults who were Facebook users scored higher on assessments of social satisfaction than those who were not [21]. Taken together, the extant literature suggests that new technologies and web-based social environments may benefit older adults, but the implementation of these environments is critical.

Objectives

Although prior research on digital health interventions has examined the effect of the peer support component using quantitative measures [11,22,23], there is great potential to employ qualitative data to better understand participants' experiences of peer support interventions and iCBT interventions for older adults. In this study, we performed a qualitative secondary analysis of data from MoodTech, an internet-based intervention for depression for adults aged 65 years and above [11]. This intervention included didactic content (in the form of lessons), interactive skills–based tools, and peer support features. The primary analysis focused on quantitative analyses of engagement metrics (participant log-ins and minutes); usability; and mental health outcomes, including depression and anxiety, and reported a reduction in depression among participants compared with those waitlisted [11].

In this study, we focused on the textual data that participants produced, either as part of the activities they engaged in along with the didactic content or in the peer-support component of the website, to understand the participants' experiences of the intervention and develop recommendations for the design of internet-based interventions for older adults. We describe three themes, namely, challenges experienced by older adults in life, difficulties experienced in the intervention, and the benefits of the online social space, using the first theme of life challenges to provide context for the other two themes.

Methods

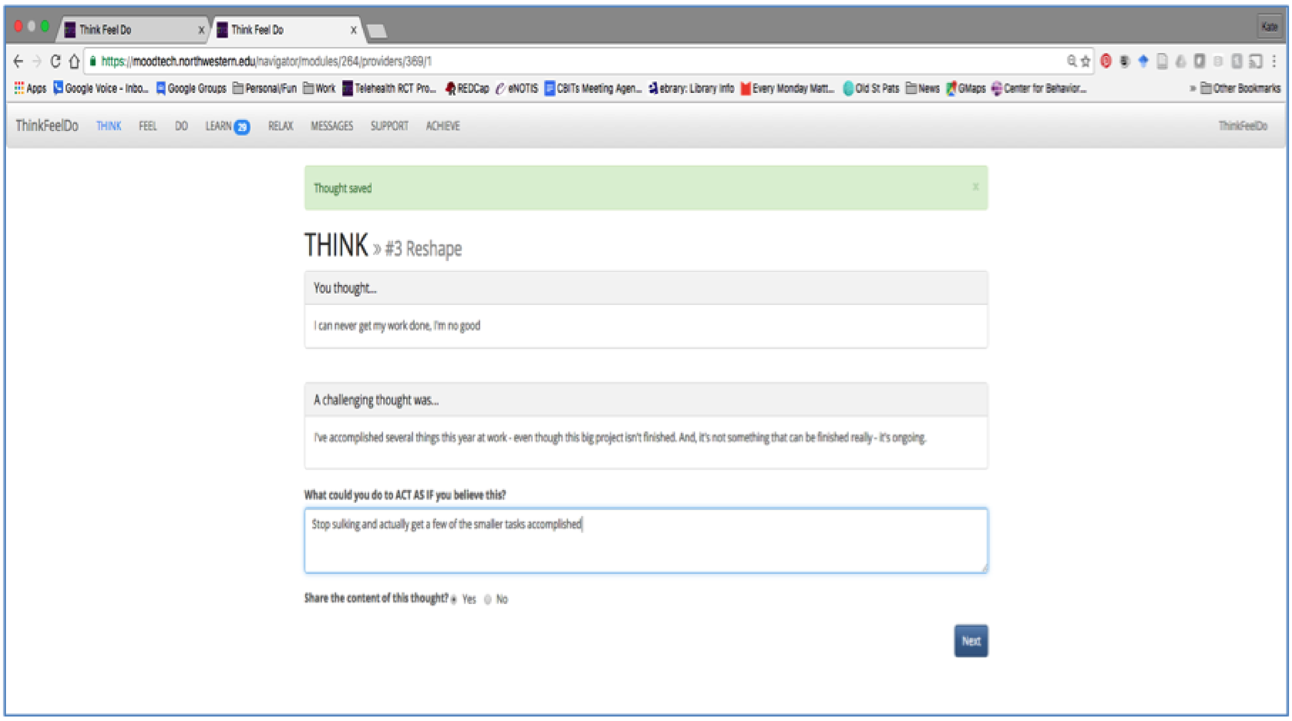
In the Methods section, we introduce the intervention, explain the data that are the focus of the analysis, and then describe the data analysis method.

MoodTech

The MoodTech study was designed to evaluate the feasibility and acceptability of an online intervention for depression based on CBT principles, developed for participants aged 65 years and above [11]. MoodTech was built on the ThinkFeelDo platform [24] and focused on five core skills: (1) cognitive restructuring (think), (2) mood and emotion monitoring (feel), (3) behavioral activation (do), (4) relaxation/mindfulness (relax), and (5) goal setting (achieve). Activities were an aspect of the *do* tool. Within the tool, participants were able to list activities completed in the last few days. Participants could plan activities to boost their mood, provide an optional prospective due date for completion, and report completion. The *think* tool asked the users to record their negative thoughts, categorize the type of cognitive distortion (eg, magnifying, minimizing and tunnel vision), reframe the thought constructively, and create an action plan (Figure 1).

The peer support version of the intervention also featured an online social environment in which the participants could interact with one another (Figure 2). This included an activity feed that automatically posted the participants' actions, including lessons read, goals set, and relaxation exercises completed. The participants could share status updates and comment on items shared by others. They also had the option of sharing thoughts they entered as part of the *think* tool and activities they completed or scheduled in the *do* tool.

Figure 1. Screenshot of harmful thoughts activity. Participants enter thoughts and create a challenging thought and action in response. The thoughts can be private or shared publicly with the peer support group, allowing peer and coach comments.



Thought saved

THINK » #3 Reshape

You thought...

I can never get my work done, I'm no good

A challenging thought was...

I've accomplished several things this year at work - even though this big project isn't finished. And, it's not something that can be finished really - it's ongoing.

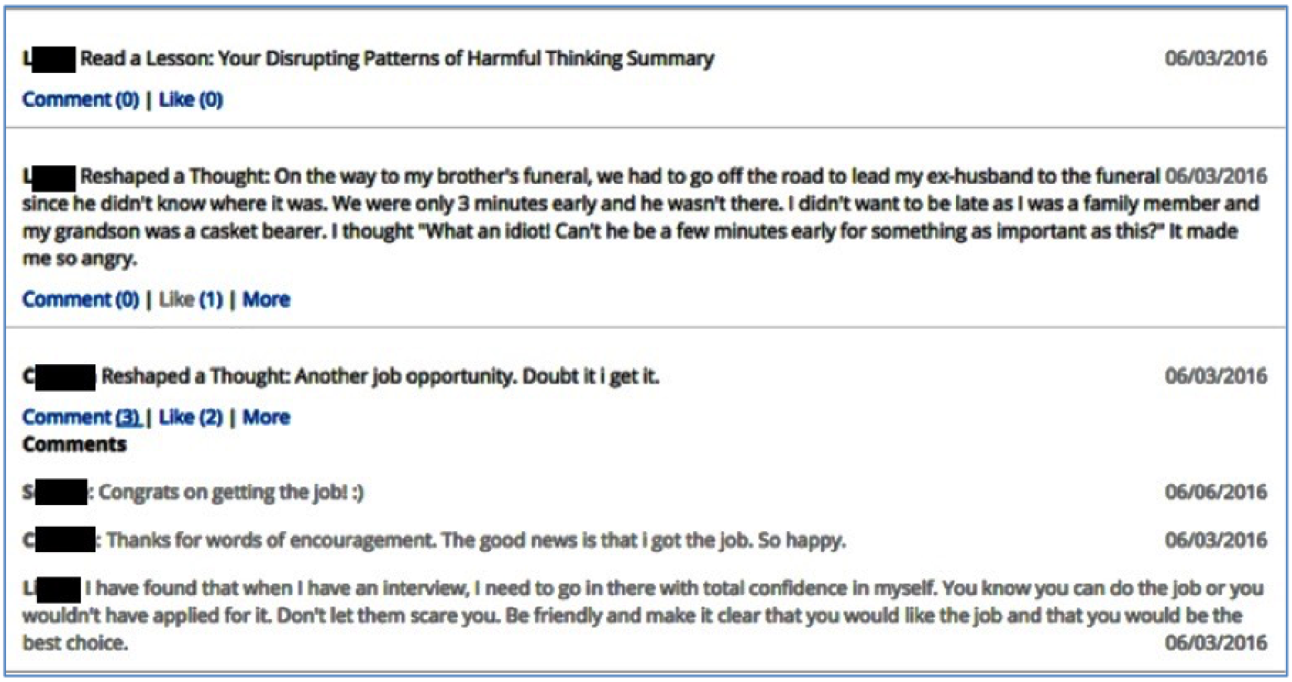
What could you do to ACT AS IF you believe this?

Stop sulking and actually get a few of the smaller tasks accomplished

Share the content of this thought? ● Yes ○ No

Next

Figure 2. Screenshot of the social peer support feature displaying a portion of social media feed where a negative thought shared publicly had 3 comments and 2 likes.



[redacted] Read a Lesson: Your Disrupting Patterns of Harmful Thinking Summary 06/03/2016

Comment (0) | Like (0)

[redacted] Reshaped a Thought: On the way to my brother's funeral, we had to go off the road to lead my ex-husband to the funeral since he didn't know where it was. We were only 3 minutes early and he wasn't there. I didn't want to be late as I was a family member and my grandson was a casket bearer. I thought "What an idiot! Can't he be a few minutes early for something as important as this?" It made me so angry. 06/03/2016

Comment (0) | Like (1) | More

[redacted] Reshaped a Thought: Another job opportunity. Doubt it i get it. 06/03/2016

Comment (3) | Like (2) | More

Comments

[redacted]: Congrats on getting the job! :) 06/06/2016

[redacted]: Thanks for words of encouragement. The good news is that i got the job. So happy. 06/03/2016

[redacted] I have found that when I have an interview, I need to go in there with total confidence in myself. You know you can do the job or you wouldn't have applied for it. Don't let them scare you. Be friendly and make it clear that you would like the job and that you would be the best choice. 06/03/2016

The study comprised 3 treatment arms, each 8 weeks in length: 1 group received the online intervention without peer support (individual internet-based intervention), another group (divided into 2 cohorts) received the online intervention with peer support (internet-based intervention with peer support), and the final group was a waitlist control group that was evaluated with no treatment during the course of intervention before being given access to the individual internet-based intervention.

Participants were primarily recruited through clinical research registries as well as online and community advertisements and clinic referrals. To participate, individuals needed to be aged ≥65 years; to have elevated depressive symptoms at screening (score of ≥8 on the patient health questionnaire-8 or a score of

>7 on the geriatric depression scale-15); to be able to read and speak English; and to have basic internet skills and access to the internet, a telephone, and a valid email address. The exclusion criteria included current engagement or planned engagement in psychotherapy, psychiatric diagnoses for which participation could be inappropriate (eg, psychotic disorder), or cognitive impairment [11].

Data Set

The MoodTech data set comprises demographics, usage, and outcomes data collected before and during the MoodTech intervention. We report the descriptive statistics and feature usage statistics for the sample.

During the intervention, participants produced textual data incidentally through their interaction with the system. First, participants produced textual data through individual activities, such as the goals they set for themselves, and their work on challenging negative thoughts. Status updates were part of the activity feed, allowing users to post unsolicited statements to their peers on any topic. Comments and *likes* could be made in response to any material in the activity feed, including thoughts or goals that were publicly shared, user profiles, or status updates, and were visible to all participants in the intervention. Comments were the only means of direct text communication among the participants and often would resemble conversation, with multiple comments back and forth attached to an original post to the feed. The participants were also able to *nudge* (a social action resembling a physical *nudge*, which carries no verbal communication) other participants to promote engagement. The intervention also included a *messages* section, where participants could interact privately with their coach. The coach could also comment on the participants' activity on the feed.

In this study, we focused on textual data from thought activities, status updates, and comments as these tended to be longer and richer, thus providing more insight into participants' experiences. The textual data were aggregated into documents by participant ID and textual data type (activities, thoughts, status posts, and comments) for analysis, such that each document represented the textual expression of 1 participant for a given data type. For example, all status posts by ID 1523 would have been aggregated into a single document. To provide context for analysis of comments, the original status post or item (eg, harmful thought) that elicited the comment was also included in these documents. We focused on the text generated as part of the *think* section or in the social space (ie, status updates and comments), which provided the richest insight into participants' experiences of the MoodTech platform and their efforts to practice the skills taught in the intervention.

Qualitative Data Analysis

Our analytic method was informed by two qualitative approaches: interpretative phenomenological analysis (IPA) [25] and grounded theory [26-28]. At the outset, the perspective with which we viewed the data was influenced by IPA, a method that places emphasis on how individuals make sense of their world, positing a close connection between what a person says or writes and his or her cognitions [25,29]. IPA is widely used in health-related studies with a focus on patient experience [29-31]. In this study, we adopted this perspective in code development, endeavoring to create codes that illustrated participants' experiences of life and the intervention.

Our analytic procedure was based on grounded theory methodology. In grounded theory, researchers code the data line by line [26]. The codes are refined through a process involving constant comparisons and constant questioning. We identified concepts of interest, namely, health, depression, aging, technology, and participation in the social networking component of the intervention, and 1 author (KS) performed line-by-line coding, creating codes that were related to one or more of these topics. Then, 2 authors (KS and AC) worked together to iteratively refine and merge codes and to create a hierarchical code structure of themes and subthemes. After KS developed the initial codes and hierarchical structure of themes, AC reviewed the codes and structure and suggested revisions, which were then reviewed by KS. The 2 authors agreed upon the final set of codes and hierarchical code structure. The codes were compared for the same participant and across participants to ensure consistency of the coding scheme, focusing on the reduction of conceptual overlap and redundancy. The two authors consulted with a clinical expert (KT) to ensure that the interpretation of the themes was consistent with the clinical knowledge of depression in older adults and with the design of the MoodTech intervention. The data analysis was performed using the qualitative data analysis software, Atlas.ti (Cleverbridge, Inc) [32].

Although we initially included activities (data from the *do* tool) in our analysis, we found that they were usually short phrases or sentences indicating what the participants planned or had already done (eg, *ate supper*, ID 1568) and thus uninformative concerning the participants' inner mental state. Therefore, we did not report the themes from the activities in this paper. In selecting the themes to report and elaborate upon, we considered the guidelines provided by Braun and Clarke [33], with frequency and conceptual relatedness of the codes to one another being key considerations. In this paper, we focused on participants' life challenges, difficulties with the program, and benefits of the program. In our selection of quotes, we endeavored to represent the diversity of voices among the sample.

The aims of our investigation were to qualitatively understand how the participants experienced the intervention material and the peer support component. With respect to the first goal, all participants, regardless of treatment arm, received the same intervention materials; thus, we analyzed the available data from all treatment arms together. With regard to the second goal of analyzing participants' experiences with the peer support component, we only used data from the one arm that received this component, the internet-based intervention with peer support.

Results

Data Set

Sample

The MoodTech study included 47 participants. Two of the original participants (1 in the individual internet-based intervention and the other in the wait list control) were not included in this data set. One participant in the individual internet-based intervention group dropped out of the study before using the intervention, and one of the 12 participants in the wait list control decided not to use the intervention after the waiting period. Most of the participants were female, white, and retired. Overall, the sample was highly educated, with most participants having a 2-year college degree or above. The summary statistics for the sample are shown in Table 1.

Table 1. Participant characteristics.

Participant characteristic	Individual internet-based intervention	Internet-based intervention with peer support	Wait list control
Age (years), mean (SD)	69.3 (3.5)	69.5 (4.3)	70.2 (4.9)
Sex, n			
Female	7	16	7
Male	4	7	4
Ethnicity, n			
Not Hispanic or Latino	11	22	11
Hispanic or Latino	0	1	0
Race, n			
Black or African American	0	1	0
White	10	19	11
More than 1 race	1	2	0
Declined to report	0	1	0
Marital status, n			
No partner	7	16	4
Partner	4	7	7
Education, n			
Some college	0	5	1
2-year college (Associate’s degree)	0	5	2
4-year college (Bachelor’s degree)	3	6	4
Master’s degree	5	5	2
Doctoral degree (PhD, MD, and JD)	3	2	2
Employment status, n			
Employed	2	3	3
Retired	9	18	7
Other	0	2	1

Feature Usage

Although most participants engaged with the features of the application they had access to, there was a great deal of variability in the extent to which the features were used (Table 2). A total of 43 out of 45 participants produced text data of at least one type. In the treatment arm with peer support, less than half of the individuals authored status updates, but almost all participants commented on at least one status post. Similarly, with nonverbal social interactions, almost all participants liked an item at least once, but only about half of the participants used the nudges. As the SDs indicate, there was substantial variability in the extent to which the participants used the features of the intervention.

Table 2. Engagement metrics for the skills and peer support components of the MoodTech intervention.

Feature used	Individual internet-based intervention (n=11)		Internet-based intervention with peer support (n=23)		Wait list control (n=11)	
	Participants, n	Participants, mean (SD)	Participants, n	Participants, mean (SD)	Participants, n	Participants, mean (SD)
Thoughts	9	9.1 (4.5)	18	11.9 (11.1)	9	10.1 (5.3)
Messages	11	11.6 (5.8)	21	9.3 (7.2)	11	10.4 (6.1)
Social interactions: verbal						
Status posts	N/A[a]	N/A	10	19.7 (16.7)	N/A	N/A
Comments	N/A	N/A	19	20.3 (23.6)	N/A	N/A
Social interactions: nonverbal						
Likes	N/A	N/A	19	25.7 (52)	N/A	N/A
Nudges	N/A	N/A	9	5.7 (8.2)	N/A	N/A

[a] N/A: not applicable.

Textual Data

Our analysis focused on the textual data produced as a part of the *think* and *do* sections of the intervention and the social space. The activities and harmful thoughts features of MoodTech were available to all 45 participants, whereas status updates and comments were only available to participants in the peer support arm (n=23). Table 3 depicts how the text data were divided into documents, with each document pertaining to a unique participant-data type pairing, and imported into Atlas.ti.

Table 3. Textual data used in qualitative data analysis.

Arm	Participants, n	Intervention features (available to all)		Peer support component features		Total documents, n
		Activities	Thoughts	Status post	Comment	
Individual internet-based intervention	11	3	11	N/A[a]	N/A	14
Internet-based intervention with peer support	23	15	20	10	19	64
Wait list control	11	1	9	N/A	N/A	10
Total	45	19	40	10	19	88

[a] N/A: not applicable.

Qualitative Data Analysis

Life Challenges

As part of the didactic component of the MoodTech intervention, participants learned how to reframe and reduce harmful thoughts. We analyzed these harmful thoughts to help us understand the challenges that participants face in life, to which they endeavored to apply the skills that they were learning. We describe these life challenges to provide context and to better understand the difficulties and benefits the participants experienced, both with the MoodTech intervention and with skill application.

Health Concerns

Participants in the MoodTech intervention expressed many challenges, including health-related problems such as chronic pain, sleep, weight, physical ability, and potentially negative results from medical tests. These often left them with a sense of regret and hopelessness:

> *I have feelings of regret and grief over my health and often wish it would speed up and not take so long until it's over. The pain levels are extremely challenging.* [ID1607, harmful thought]
>
> *When I'm in pain, I feel as though it's going to last forever, that I'll never get better...* [ID1816, harmful thought]

Other Challenges

Other than health-related challenges, participants also experienced personal challenges. These could be intertwined with their recollections of the past and concerns about the future. Family played a prominent role. Participants worried about the welfare of their children and their parents, were fearful of losing loved ones, and wrote of tensions they experienced interacting with other family members. They were also concerned about financial status, professional status, and legacy:

> *I was not a good mother, my kids had to do without so much and work harder.* [ID1651, harmful thought]
>
> *I am afraid that I will not have enough money.* [ID1596, harmful thought]

Participants also had concerns about how they were perhaps perceived by others. There were various reasons for this, including feeling self-conscious about physical appearance, but some participants were also concerned about how they were perceived in terms of intellect. Participants expressed concern about cognitive difficulties and memory issues:

> *I can't get my thoughts together; I'm getting worse.* [ID1885, harmful thought]
>
> *I wish my memory was as it once was. I am ALWAYS forgetting names, words, and where I put things as three examples. I feel embarrassed and diminished in the eyes of my family.* [ID1838, harmful thought]

Difficulties That Participants Experienced in the Intervention

Technical Difficulties Leading to Self-Doubts

Participants experienced various difficulties in working with the MoodTech intervention. First, in terms of the technology, participants often encountered issues with the system, such as creating a text entry, not having the entry saved, and then having to reenter it. These issues led to confusion and frustration, which were accompanied by the participants questioning their memory and abilities:

> *The program confuses me and adds to my litany of reasons I think I am a dunce!* [ID1523, comment]
>
> *I think the software challenges on site make several of us feel like dummies and that is no help to alleviate depression.* [ID1607, harmful thought]

Being Overwhelmed by the Pace of the Intervention

Participants felt overwhelmed by the amount of work they perceived that they had as well as the pace of the intervention:

> *How will I ever get all of this work done in just 7 weeks that I have left?* [ID1599, harmful thought]
>
> *What's the matter with me that I have a hard time keeping up with the lessons in this program and not working on them as much as I should* [ID1781, harmful thought]
>
> *Anyone else finding this program overwhelming. Todays lesson is good but difficult.* [ID1596, status post]
>
> *I'm feeling overwhelmed too. The weekly/daily pressures are too fast-paced for me. I suppose the program needs a schedule but it doesn't take into account the actual pace and work of being depressed. I feel behind too.* [ID1550, comment]

The problems that participants experienced with the program compounded insecurities and anxieties that participants already had, which led to an increased sense of pressure and frustration.

The Benefits of Social Space

Despite the difficulties that the participants experienced with MoodTech, they also experienced positive interactions in the peer support space. These included affirmation and encouragement from peers and coaches, collaborative efforts at problem solving, and positive affective experiences from the social experience of the website.

Empathy and Encouragement

Participants expressed empathy and provided encouragement to one another for the difficulties they experienced in life:

> *This is just a temporary setback (having to keep off your feet for most of the week). You can do it (lost weight) – don't give up trying. Mind over matter!* [ID1816, comment]
>
> *Looks like you're on top of the problem and are looking ahead of what comes next.* [ID1813, comment]

Appreciation and Affirmation

At the outset, participants expressed appreciation for and affirmed the contributions of others on the website:

> *This is a very good insight, [ID 1607]. Although I have not posted very frequently, I feel that you have been very courageous and open during this study.* [ID1596, comment]
>
> *Dear [ID 1781], I have been thinking about your post since last week but couldn't put my feelings into words. Just wanted you to know that I admire your honesty and the courage that it must have taken to share this post.* [ID1760, comment]

Collaborative Problem Solving

Participants also engaged in collaborative discussions about the problems they were trying to address as part of the MoodTech intervention. For example, in the following exchange, 2 participants discussed the nature of their tendency to worry:

> *I worry too much – guess I feel that by worrying about all the contingencies, I can somehow control one or two of them. Anyone else have that problem...* [ID1816, status post]
>
> *I also struggle with worrying a lot. [ID 1816] ...I think you nailed it when you said that the reason is that perhaps by worrying, you can control one or two of them...* [ID1723, comment]

Participants also helped one another figure out how to utilize the website. For example, several participants were confused about how to go from the email notifications of a new comment on their post to the comments themselves. They were able to explain to one another in the social space how to navigate to the particular content that they wanted:

> *Until I figured out to go to my profile page & scroll down, I'd get the email saying someone commented to me then I'd go to the main page & scroll down through ALL comments until I found mine.* [ID1550, comment]
>
> *I just got a couple of notes that someone “”LIKED“” a couple of my comments. But it's hard to figure out who liked what -- can't I get a link directly to what was liked, so I can acknowledge it?* [ID1816, status post]
>
> *It was confusing to me at first, also. Scroll down all the things that you posted on the Feed. When you find one with a number in parenthesis behind the comment*

> *or like, click on the number. That will take you to the comment and who liked it or made a comment. HOpe this helps.* [ID1723, comment]

Discussion

Principal Findings

In this paper, we present a qualitative analysis of textual data collected from MoodTech, a pilot study of an internet-based intervention with a peer support component for older adults with symptoms of depression. We analyzed 3 main types of data from the intervention: harmful thoughts, status updates, and comments. We present the results through 3 main themes: (1) challenges that participants experienced in life, (2) difficulties that participants experienced during the course of the intervention, and (3) the benefits that they derived from participating in the social space. The thoughts section of the data (individual) provided the most insight into difficulties experienced in the intervention, whereas the status post and comment sections (peer support) provided insight into the challenges and difficulties experienced as well as the benefits of having a peer support component in an intervention.

At the outset, participants shared various life challenges that they experienced, including health concerns, self-consciousness, recollections of the past, and anxiety about the future. These concerns provide context for the difficulties that the participants experienced as they engaged in the intervention activities, including being overwhelmed by the pace of the intervention and technical issues. The difficulties they experienced with the intervention fueled the insecurities they were experiencing in life. However, the peer support component appeared to benefit the participants by providing a medium for empathy, encouragement, appreciation, and affirmation, and collaborative problem solving.

The life challenges mentioned by the participants in this study went beyond health-related challenges to include other challenges in which the effects of time and aging were clearly related. This characterization of life challenge resonates with extant literature on aging, which has noted that there is a difference between objective (health-related) and subjective perceptions of health [34]; people with multiple chronic conditions may rate themselves as healthy and active, whereas those with few or no conditions may rate their health more poorly, because of psychological and social factors, coping mechanisms, and resilience. The data collected from the social space can provide insights for researchers into the challenges experienced by the participants in the aging process as well as about how these challenges may affect the participants' experiences of the intervention.

Social Space for Information Sharing and Mutual Encouragement

The MoodTech intervention not only offered participants an opportunity to practice the skills that they were learning individually but also with the option of sharing them with others. Although the participants in this study experienced frustration and confusion due to technical difficulties, the participants' written utterances also suggested that the ability to interact in the social space was a positive experience and that they valued the support and camaraderie of others.

The interactions among the participants in the social space showed that the participants were motivated to help one another with technical issues as well as with the difficulties they experienced in life or with the intervention content. The content and shared understanding from peer interactions made it possible to brainstorm coping strategies or at least take comfort in knowing that others shared their experiences. With regard to the technology, participants understood the frustrations that others were experiencing as well as the insecurities that were perhaps triggered by age-related factors and were quick to offer advice and workarounds to reduce others' struggles. The findings thus illustrate that a social networking component within the context of a cognitive behavioral intervention can perhaps provide a synergistic contribution to the intervention and enrich and deepen the learning that participants are engaged in with the educational modules.

One question that might arise from this is if there was an optimal level of usage of the features. There was extensive variability in the extent to which the study features were used. Although not all participants posted status updates, almost all participants commented at least once and liked the items shared by others. Participants commented on others' thoughts and behaviors and expressed admiration for the courage of others who shared their experiences.

Only about half of the participants posted status updates and nudges, and we can only speculate about the causes. Although the reasons for lower usage of these features are unclear, perhaps posting a status update required more courage or effort than simply commenting on the posts of others. With regard to nudges, some possible reasons might be that the meaning of a nudge was less intuitive to participants or more personal than they wanted it to be.

Overall, these findings are consistent with prior literature, which reported that sharing of experiences on online support groups can lead to improved sentiment [35] and patient empowerment [36]. Of particular interest perhaps are contagion effects resulting from participation in a group discussion. As described in the Results, we observed various positive effects of group interactions such as being inspired by the actions of others and resonance in terms of shared experiences leading to subsequent sharing of similar experiences. In some cases, the shared experiences were related to difficulties with the intervention. One participant expressing feelings of frustration with the program can lead to others sharing similar feelings. However, the expression of negative sentiments is not necessarily undesirable. Expressing frustration gives participants a chance to release these tensions and frustrations so that they can be overcome, and, in this intervention, it also led to collaborative problem solving.

In this study, some participants were clearly more active than others, and participants also differed in terms of the nature of content they contributed. Previous research has found that members of online communities play different roles [37]. Previous research investigating lurkers (those who read but do not post) and posters has reported that posters tend to report having their expectations met more often than lurkers [38], those who received emotional support in an online support group were less likely to drop out [39], and online support group members provide different forms of support to one another [40].

On the basis of the comments made in the social space, we suspect that some participants in this study did find the content of others useful even if they did not contribute content themselves and that many perhaps had respect for those who had the courage to do so. It is likely that some participants, whether they contributed or not, benefited from the social space, in terms of both informational and social interactions. However, more research is needed to better understand how different forms of engagement with the social space may be associated with health-related outcomes.

Design Implications

Participants in this study experienced a substantial amount of frustration due to technical issues. In some but not all cases, this appeared to be exacerbated by other problems they experienced, such as cognitive issues. The problems they experienced led to uncertainty about if they or the system were the cause of the problem. Although the technical difficulties faced by the participants were those that could occur with any system, considering them in the context of age-related factors is critical for delivering effective treatment.

The idea that technological interventions for older adults should consider age-related factors is not new. Prior research, for example, has recommended that technology-based training for older adults should be well-structured, provide feedback and adaptive guidance, include metacognitive prompts, include a user interface that is simple and consistent, and take into consideration extant knowledge of cognitive load [41]. This can be difficult for complex multicomponent educational interventions.

This study provides insights into the design strategies for minimizing confusion and frustration on the part of older adults in the design of educational interventions. At the outset, one of the problems we observed was that participants felt mounting pressure due to the perception that they could not finish their assignments. With MoodTech, a potential alternative design might be to present personalized educational objectives to participants, paced to suit the needs and goals of the participants over time. For example, participants could be told that MoodTech would provide assignments that depended on how they were doing, in a manner similar to how an adaptive system would work. This way, participants would not feel despair at not being able to finish the modules. This might be more difficult in the setting of a peer-supported intervention in which participants are aiming to move through the material as a group. Presenting content in an adaptive fashion might lead to some participants referring to content that others might not have seen. Thus, there is a design tension that must be considered when providing a more personalized experience and a unified group experience that members share.

It is useful to consider the multiple factors that may be at play in the perception of pressure to finish the assignments. Lack of motivation, feeling inadequate, and feeling pressured can also be symptoms of depression, which the intervention is intended to address; it may not be in the participants’ interests to cater to their preferences completely. The design of digital mental health technologies involves a competing tension between promoting engagement and achieving improvements in symptoms and well-being. Behavior change often involves confronting problems that can, at times, be uncomfortable. Thus, the goal of addressing the issues raised in this study may be to minimize the pressure experienced by users by allowing meaningful engagement, recognizing that some users may nevertheless continue to experience some discomfort.

A second design takeaway would be to simplify the instructions and training that are provided to the participants. Previous research has also argued that the availability of support is critical as older adults are learning a technology, although this support can take various forms, including step-by-step guidance, an error-friendly space where someone is available to fix problems if they arise, and manuals [14]. Owing to the complexity of the system used in this study, the training materials that the participants received were extensive, and it may have been difficult to find what they needed at the appropriate time. Providing simpler training materials, a short summary of key actions that is accessible from anywhere in the program, and training videos could perhaps enable participants to engage with the site more easily and ameliorate the confusion and frustration experienced during the study. Having these features might also reduce the burden on coaches for providing technical support and training and preventing intervention dropout.

Although disorientation and cognitive overload are well-recognized as concerns in the interface design [42], this qualitative analysis has shown that they take a particular meaning with this population and that addressing these issues is paramount. User interface adaptations that involve the reduction of interface complexity or enhancement of assistance features can be particularly important for older adults to mitigate the effects of cognitive decline and vision loss [43] and prevent normal user interface difficulties from being interpreted as evidence of cognitive decline or lack of competence.

Limitations and Future Directions

Our analysis had various limitations that could also serve as potentially fruitful directions for subsequent research. First, one of the data types that we analyzed, reframing negative thoughts, were a skill that participants were explicitly directed to practice. As such, these thoughts were inherently negative in nature, and both the life challenges and negative aspects of the participants’ intervention experiences might be more apparent than positive aspects due to our selection of data to analyze. Second, in addition to the quantitative and qualitative analyses that have been conducted, analytical methods that model the participants’ interactions, such as network analysis, could have provided additional insight into how the participants interacted with one another and could be a fruitful direction for future research. Third, the findings are based on a secondary analysis of the content produced during the course of the intervention; we did not directly conduct an exit interview asking the participants to comment on their experiences or specific features such as nudges, which could have led to additional insights to inform the design and features of the intervention. Finally, our sample primarily comprised individuals who were female, white, and highly educated. There is a need to further explore how individuals from other backgrounds might experience this intervention.

Conclusions

In this paper, we analyzed textual data from an internet intervention with a peer support component for older adults with symptoms of depression. The textual data provided insight into the participants' experiences with the intervention as well as design considerations for developing complex technological interventions that support the challenges that participants may experience due to aging and cognitive difficulties. First, technical issues encountered by older adults within the context of the intervention can interact with and exacerbate insecurities that they may experience in life, and it is important to consider how intervention components might be designed to mitigate these issues. Second, peer support can be employed as a mechanism to facilitate communication, support, and collaborative problem solving among participants in an intervention. It is our hope that insights from this study inform the design of other iCBT interventions for older adults.

Acknowledgments

This project was supported by the National Institute of Mental Health (grants P20 MH090318 and R01 MH095753; PI: DM). Data collection used REDCap electronic data tools supported, in part, by the National Institutes of Health National Center for Advancing Translational Sciences, grant UL1TR001422.

Authors' Contributions

AC, KS, and KT were involved in the conceptualization of this secondary data analysis. CR and LB deidentified the data that was used in the analysis. AC and KS analyzed the data, and KT provided consultation in the analysis. AC and KS drafted the manuscript, and KT, CR, LB, and DM provided feedback on multiple versions of the manuscript. All authors gave permission for the final version to be published.

Conflicts of Interest

DM has accepted consulting fees from Apple Inc, Pear Therapeutics, Otsuka Pharmaceutical Inc, and has an ownership interest in Adaptive Health, Inc.

References

1. Byers AL, Yaffe K, Covinsky KE, Friedman MB, Bruce ML. High occurrence of mood and anxiety disorders among older adults: the national comorbidity survey replication. Arch Gen Psychiatry 2010 May;67(5):489-496 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.35] [Medline: 20439830]
2. Kok RM, Reynolds CF. Management of depression in older adults: a review. J Am Med Assoc 2017 May 23;317(20):2114-2122. [doi: 10.1001/jama.2017.5706] [Medline: 28535241]
3. Lépine JP, Briley M. The increasing burden of depression. Neuropsychiatr Dis Treat 2011;7(Suppl 1):3-7 [FREE Full text] [doi: 10.2147/NDT.S19617] [Medline: 21750622]
4. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the national comorbidity survey replication. Arch Gen Psychiatry 2005 Jun;62(6):629-640. [doi: 10.1001/archpsyc.62.6.629] [Medline: 15939840]
5. Fiske A, Wetherell JL, Gatz M. Depression in older adults. Annu Rev Clin Psychol 2009;5:363-389 [FREE Full text] [doi: 10.1146/annurev.clinpsy.032408.153621] [Medline: 19327033]
6. Jonsson U, Bertilsson G, Allard P, Gyllensvärd H, Söderlund A, Tham A, et al. Psychological treatment of depression in people aged 65 years and over: a systematic review of efficacy, safety, and cost-effectiveness. PLoS One 2016;11(8):e0160859 [FREE Full text] [doi: 10.1371/journal.pone.0160859] [Medline: 27537217]
7. Xiang X, Wu S, Zuverink A, Tomasino KN, An R, Himle JA. Internet-delivered cognitive behavioral therapies for late-life depressive symptoms: a systematic review and meta-analysis. Aging Ment Health 2019 Mar 27:1-11 epub ahead of print. [doi: 10.1080/13607863.2019.1590309] [Medline: 30913898]
8. Ballantyne A, Trenwith L, Zubrinich S, Corlis M. I feel less lonely: what older people say about participating in a social networking website. Qual Ageing Older Adults 2010 Sep 13;11(3):25-35. [doi: 10.5042/qiaoa.2010.0526]
9. Khosravi P, Rezvani A, Wiewiora A. The impact of technology on older adults' social isolation. Comput Hum Behav 2016 Oct;63:594-603. [doi: 10.1016/j.chb.2016.05.092]
10. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston L, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided-self-guided internet-delivered cognitive-behavioural therapy. BJPsych Open 2016 Jan;2(1):50-58 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002139] [Medline: 27703754]
11. Tomasino KN, Lattie EG, Ho J, Palac HL, Kaiser SM, Mohr DC. Harnessing peer support in an online intervention for older adults with depression. Am J Geriatr Psychiatry 2017 Oct;25(10):1109-1119 [FREE Full text] [doi: 10.1016/j.jagp.2017.04.015] [Medline: 28571785]

12. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
13. Choi M, Kong S, Jung D. Computer and internet interventions for loneliness and depression in older adults: a meta-analysis. Healthc Inform Res 2012 Sep;18(3):191-198 [FREE Full text] [doi: 10.4258/hir.2012.18.3.191] [Medline: 23115742]
14. Barnard Y, Bradley MD, Hodgson F, Lloyd AD. Learning to use new technologies by older adults: perceived difficulties, experimentation behaviour and usability. Comput Hum Behav 2013 Jul;29(4):1715-1724. [doi: 10.1016/j.chb.2013.02.006]
15. Hill R, Betts LR, Gardner SE. Older adults' experiences and perceptions of digital technology: (dis)empowerment, wellbeing, and inclusion. Comput Hum Behav 2015 Jul;48:415-423. [doi: 10.1016/j.chb.2015.01.062]
16. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
17. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 1;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
18. Chiu C, Hu Y, Lin D, Chang F, Chang C, Lai C. The attitudes, impact, and learning needs of older adults using apps on touchscreen mobile devices: results from a pilot study. Comput Hum Behav 2016 Oct;63:189-197. [doi: 10.1016/j.chb.2016.05.020]
19. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among older adults. Comput Hum Behav 2012 Mar;28(2):496-499. [doi: 10.1016/j.chb.2011.10.021]
20. Braun MT. Obstacles to social networking website use among older adults. Comput Hum Behav 2013 May;29(3):673-680. [doi: 10.1016/j.chb.2012.12.004]
21. Bell C, Fausset C, Farmer S, Nguyen J, Harley L, Fain W. Examining Social Media Use Among Older Adults. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. 2013 Presented at: HT'13; May 1-3, 2013; Paris, France p. 158-163. [doi: 10.1145/2481492.2481509]
22. Nahm ES, Barker B, Resnick B, Covington B, Magaziner J, Brennan PF. Effects of a social cognitive theory-based hip fracture prevention web site for older adults. Comput Inform Nurs 2010;28(6):371-379 [FREE Full text] [doi: 10.1097/NCN.0b013e3181f69d73] [Medline: 20978408]
23. Proudfoot J, Parker G, Manicavasagar V, Hadzi-Pavlovic D, Whitton A, Nicholas J, et al. Effects of adjunctive peer support on perceptions of illness control and understanding in an online psychoeducation program for bipolar disorder: a randomised controlled trial. J Affect Disord 2012 Dec 15;142(1-3):98-105. [doi: 10.1016/j.jad.2012.04.007] [Medline: 22858215]
24. Schueller S, Mohr D. European Union Digital Library. 2015. Initial Field Trial of a Coach-Supported Web-Based Depression Treatment URL: http://eudl.eu/doi/10.4108/icst.pervasivehealth.2015.260115 [accessed 2017-08-24]
25. Smith JA, Osborn M. Interpretative phenomenological analysis. In: Smith JA, editor. Qualitative Psychology: A Practical Guide To Research Methods. London, UK: Sage Publications; 2008:53-80.
26. Charmaz K. 'Discovering' chronic illness: using grounded theory. Soc Sci Med 1990;30(11):1161-1172. [doi: 10.1016/0277-9536(90)90256-r] [Medline: 2360052]
27. Charmaz K. Constructing Grounded Theory. London, UK: Sage Publications; 2014.
28. Glaser B, Strauss A. The Discovery Of Grounded Theory: Strategies For Qualitative Research. Chicago, IL: Aldine Transaction; 1967.
29. Knight MT, Wykes T, Hayward P. 'People don't understand': an investigation of stigma in schizophrenia using interpretative phenomenological analysis (IPA). J Ment Health 2009 Jul 6;12(3):209-222. [doi: 10.1080/0963823031000118203]
30. Osborn M, Smith J. The personal experience of chronic benign lower back pain: an interpretative phenomenological analysis. Br J Health Psychol 1998;3(1):65-83. [doi: 10.1111/j.2044-8287.1998.tb00556.x]
31. Mulveen R, Hepworth J. An interpretative phenomenological analysis of participation in a pro-anorexia internet site and its relationship with disordered eating. J Health Psychol 2006 Mar;11(2):283-296. [doi: 10.1177/1359105306061187] [Medline: 16464925]
32. Hwang S. Utilizing qualitative data analysis software: A Review of Atlas.ti. Soc Sci Comput Rev 2007 Dec 18;26(4):519-527. [doi: 10.1177/0894439307312485]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Tkatch R, Musich S, MacLeod S, Kraemer S, Hawkins K, Wicker ER, et al. A qualitative study to examine older adults' perceptions of health: keys to aging successfully. Geriatr Nurs 2017;38(6):485-490 [FREE Full text] [doi: 10.1016/j.gerinurse.2017.02.009] [Medline: 28341064]
35. Qiu B, Zhao K, Mitra P, Wu D, Caragea C, Yen J, et al. Get Online Support, Feel Better–Sentiment Analysis and Dynamics in an Online Cancer Survivor Community. In: Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. 2011 Presented at: PASSAT'11; October 9-11, 2011; Boston, MA, USA p. 274-281 URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6113125 [doi: 10.1109/PASSAT/SocialCom.2011.127]

36. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
37. Jones R, Sharkey S, Smithson J, Ford T, Emmens T, Hewis E, et al. Using metrics to describe the participative stances of members within discussion forums. J Med Internet Res 2011 Jan 10;13(1):e3 [FREE Full text] [doi: 10.2196/jmir.1591] [Medline: 21239373]
38. Nonnecke B, Preece J, Andrews D. What Lurkers and Posters Think of Each Other. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences. 2004 Presented at: HICSS'04; January 5-8, 2004; Big Island, HI, USA URL: http://ieeexplore.ieee.org/abstract/document/1265462/ [doi: 10.1109/hicss.2004.1265462]
39. Wang Y, Kraut R, Levine J. To Stay or Leave?: The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW'12; February 11-15, 2012; Seattle, WA URL: http://dl.acm.org/citation.cfm?id=2145329 [doi: 10.1145/2145204.2145329]
40. Pfeil U, Zaphiris P. Investigating social network patterns within an empathic online community for older people. Comput Hum Behav 2009 Sep;25(5):1139-1155. [doi: 10.1016/j.chb.2009.05.001]
41. Wolfson NE, Cavanagh TM, Kraiger K. Older adults and technology-based instruction: optimizing learning outcomes and transfer. Acad Manag Learn Educ 2014 Mar;13(1):26-44. [doi: 10.5465/amle.2012.0056]
42. George Saadé R, Alexandre Otrakji C. First impressions last a lifetime: effect of interface type on disorientation and cognitive load. Comput Hum Behav 2007 Jan;23(1):525-535. [doi: 10.1016/j.chb.2004.10.035]
43. Machado E, Singh D, Cruciani F, Chen L, Hanke S, Salvago F, et al. A Conceptual Framework for Adaptive User Interfaces for Older Adults. In: Proceedings of the International Conference on Pervasive Computing and Communications Workshops. 2018 Presented at: PERCOMW'18; March 19-23, 2018; Athens, Greece. [doi: 10.1109/percomw.2018.8480407]

Abbreviations

CBT: cognitive behavioral therapy
iCBT: internet-based cognitive behavioral therapy
IPA: interpretative phenomenological analysis

Edited by G Eysenbach; submitted 23.12.19; peer-reviewed by D Becker, N Lau; comments to author 07.03.20; accepted 23.03.20; published 16.06.20.

Please cite as:
Chen AT, Slattery K, Tomasino KN, Rubanovich CK, Bardsley LR, Mohr DC
Challenges and Benefits of an Internet-Based Intervention With a Peer Support Component for Older Adults With Depression: Qualitative Analysis of Textual Data
J Med Internet Res 2020;22(6):e17586
URL: https://www.jmir.org/2020/6/e17586
doi:10.2196/17586
PMID:32543448

Original Paper

Evaluation of an On-Demand Mental Health System for Depression Symptoms: Retrospective Observational Study

Sarah Kunkle[1], MPH; Manny Yip[1], MSc; Watson Ξ[1], PhD; Justin Hunt[1], MD

Ginger, San Francisco, CA, United States

Corresponding Author:
Sarah Kunkle, MPH
Ginger
116 New Montgomery St Suite 500
San Francisco, CA, 94105
United States
Phone: 1 7175197355
Email: sarah@ginger.io

Abstract

Background: Depression is an extremely prevalent issue in the United States, with an estimated 7% of adults experiencing at least one major depressive episode in 2017. Although psychotherapy and medication management are effective treatments for depression, significant barriers in accessing care persist. Virtual care can potentially address some of these obstacles.

Objective: We conducted a preliminary investigation of utilization characteristics and effectiveness of an on-demand health system for reducing depression symptoms.

Methods: Data were analyzed from 1662 users of an on-demand mental health system that includes behavioral health coaching, clinical services (therapy and psychiatry), and self-guided content and assessments primarily via a mobile app platform. Measures included engagement characterized by mobile app data, member satisfaction scores collected via in-app surveys, and depression symptoms via the Patient Health Questionnaire-2 (PHQ-2) at baseline and 8-12 week follow-up. Descriptive statistics are reported for measures, and pre/post-PHQ-2 data were analyzed using the McNemar test. A chi-square test was used to test the association between the proportion of individuals with an improvement in PHQ-2 result and care modality (coaching, therapy, and psychiatry, or hybrid).

Results: During the study period, 65.5% of individuals (1089/1662) engaged only in coaching services, 27.6% of individuals (459/1662) were engaged in both coaching and clinical services, 3.3% of individuals (54/1662) engaged only in clinical services, and 3.7% of individuals (61/1662) only used the app. Of the 1662 individuals who completed the PHQ-2 survey, 772 (46.5%) were considered a positive screen at intake, and 890 (53.6%) were considered a negative screen at intake. At follow-up, 477 (28.7%) of individuals screened positive, and 1185 (71.3%) screened negative. A McNemar test showed that there was a statistically significant decrease in the proportion of users experiencing depressed mood and anhedonia more than half the time at follow-up (*P*<.001). A chi-square test showed there was no significant association between care modality and the proportion of individuals with an improvement in PHQ-2 score.

Conclusions: This study provides preliminary insights into which aspects of an on-demand mental health system members are utilizing and levels of engagement and satisfaction over an 8-12 week window. Additionally, there is some signal that this system may be useful for reducing depression symptoms in users over this period. Additional research is required, given the study limitations, and future research directions are discussed.

(J Med Internet Res 2020;22(6):e17902) doi:10.2196/17902

KEYWORDS

mental health; depression; digital health; therapy; coaching; behavioral health; virtual care

Introduction

Depression is one of the most prevalent and impactful health conditions in the United States, according to multiple data sources. Conservative estimates from national surveys show that 7%-8% of American adults are affected by major depression annually [1,2]. A national health index of more than 41 million people, compiled by one of the largest health insurance providers, ranked depression second only to hypertension in its impact on American longevity and quality of life [3]. In addition to the direct symptoms caused by depression, the majority of adults with depression report at least some difficulty with work, home, or social activities due to their symptoms [2]. Thus, the health burden of depression translates into a significant social and economic burden [4].

Existing Approaches and Barriers

Although there are effective treatments (including pharmacological and psychotherapy interventions) for depression and other mental health conditions, the current US healthcare system does not adequately address the many social, medical, and economic barriers. Access to formal behavioral health treatment remains a challenge because of a complex collection of barriers, including perceived public or social stigma, inadequate behavioral health workforce in certain geographic regions, and poor insurance coverage and related financial costs [5].

The mental health workforce shortage is more acute for some geographic regions; nearly 40 percent of Americans live in areas designated by the federal government as having a shortage of mental health professionals [6,7]. Access challenges include general barriers such as high cost, time, and transportation, but also specific obstacles like stigma and treatment preference [8]. Having access to a preferred choice of treatment improves treatment initiation, adherence, and outcomes [9,10].

Collaborative Care Model

The primary care-based collaborative care model augments the model of behavioral health care being provided in primary care and replaces it with a three-pronged team, including the primary care physician, a care manager, and a psychiatric consultant. This model helps to spread the expertise of the psychiatric consultant across a population of primary care patients through the actions of the care manager and the prescribing of the primary care provider (PCP); therefore, a larger population of primary care patients can receive evidence-based mental health care [11].

While the collaborative care model has a strong evidence base and is beginning to receive recognition from large payors such as Centers for Medicare & Medicaid Services (CMS), challenges remain in the real-world implementation of the model. In primary care settings, it remains difficult for overburdened PCPs to prescribe psychotropic medications in a 10-15 minute encounter that can often be focused on many chronic health conditions. It can also be challenging for a PCP to digest the recommendations of the care manager or psychiatric consultant team [12].

Beyond the challenges of limited time and competing priorities within a brief appointment, the collaborative care model only addresses those patients who present to primary care with behavioral health symptoms that rise to the level of a diagnosed DSM-V mental illness, such as Major Depression or Generalized Anxiety Disorder, or the symptoms are severe enough to trigger a full PHQ-9 or GAD-7. There is another “hidden” population of individuals suffering from upstream subclinical stress and behavioral health symptoms who do not present to a PCP because they do not have a comorbid medical problem, or they may not have easy geographic or financial access to a regular PCP. In fact, according to a 2018 Kaiser poll, many Millenials report not having and not even desiring a regular PCP [13].

Telehealth and Digital Interventions

The US healthcare system has attempted to close this chasm of unmet behavioral health needs and concurrent poor quality of care through a variety of different mechanisms, including telemedicine and other digital solutions. Telehealth has become the standard to address geographic barriers to care in rural regions over nearly three decades in systems such as the Department of Veteran Affairs, where evidence-based psychotherapy and medication management can be delivered at an equivalent level of quality as in-person care [14]. A 2016 systematic review showed that telephone-based interventions had promised in reducing symptoms of depression and anxiety [15]. Unfortunately, most telemental health models primarily address the geographic maldistribution of providers as opposed to the sheer lack of full-time providers across the country.

Smartphone-based treatments have also shown promise in managing depression according to a 2017 meta-analysis [16]. Given the heterogeneity of these interventions, many studies highlight the need to establish which aspects of these technologies are most active for a given population. Thus, characterizing the features, engagement, and users of specific products can help enhance our understanding of how these new technologies and systems work for different populations.

In the broader digital health landscape, there have been calls for companies to produce and publish evidence on outcomes, including engagement and clinical outcomes [17]. There are multiple dimensions for evaluating these types of products, and many organizations, including the US FDA, the UK NHS, the APA, have proposed frameworks to guide informed decision making and evaluation. Common categories of evaluation include privacy and security, evidence base, ease of use, and data integration [18].

This exploratory study aimed to investigate the initial effectiveness of a novel on-demand system by describing utilization and satisfaction measures in addition to evaluating changes in self-reported depression symptoms. Previous research has proposed some guidelines for conducting studies in real-world settings in contrast to highly controlled efficacy studies, including observing utilization, satisfaction, and outcomes in actual practice [19]. This is consistent with principles of implementation research, which seeks to understand intervention in real-world conditions, rather than trying to control for conditions or to remove their influence as causal effects [20]. Studies of other digital mental health technologies have used this approach to investigate initial effectiveness in addition to guiding future research directions [21,22].

Methods

Participants

Individuals in this study are employees or health plan members who have access to the Ginger system as part of their employer or health plan benefits. The Ginger system includes the following exclusionary criteria where telehealth is likely not appropriate:

1. Active suicidal ideation
2. Active high-risk self-harm behavior
3. Two or more hospitalizations within the past 6 months for psychiatric reasons
4. Certain symptoms of psychosis that are poorly managed (eg, the member is not med compliant, or symptoms are unresponsive to treatment), likely incompatible with telehealth
5. A primary diagnosis of a substance use disorder, or moderate to severe substance abuse issues, due to the high complexity, severity, and risk frequently associated with such members, as well as the need for specialized care
6. Active eating disorders with symptoms considered to be high risk
7. Ongoing grave disability, including bipolar patients with active mania/hypomania or mixed episodes who are unmedicated or who have poor compliance with medication regimen over time
8. Two or more medical hospitalizations in the last month due to the high likelihood that the individual has a poorly controlled medical condition that requires close monitoring

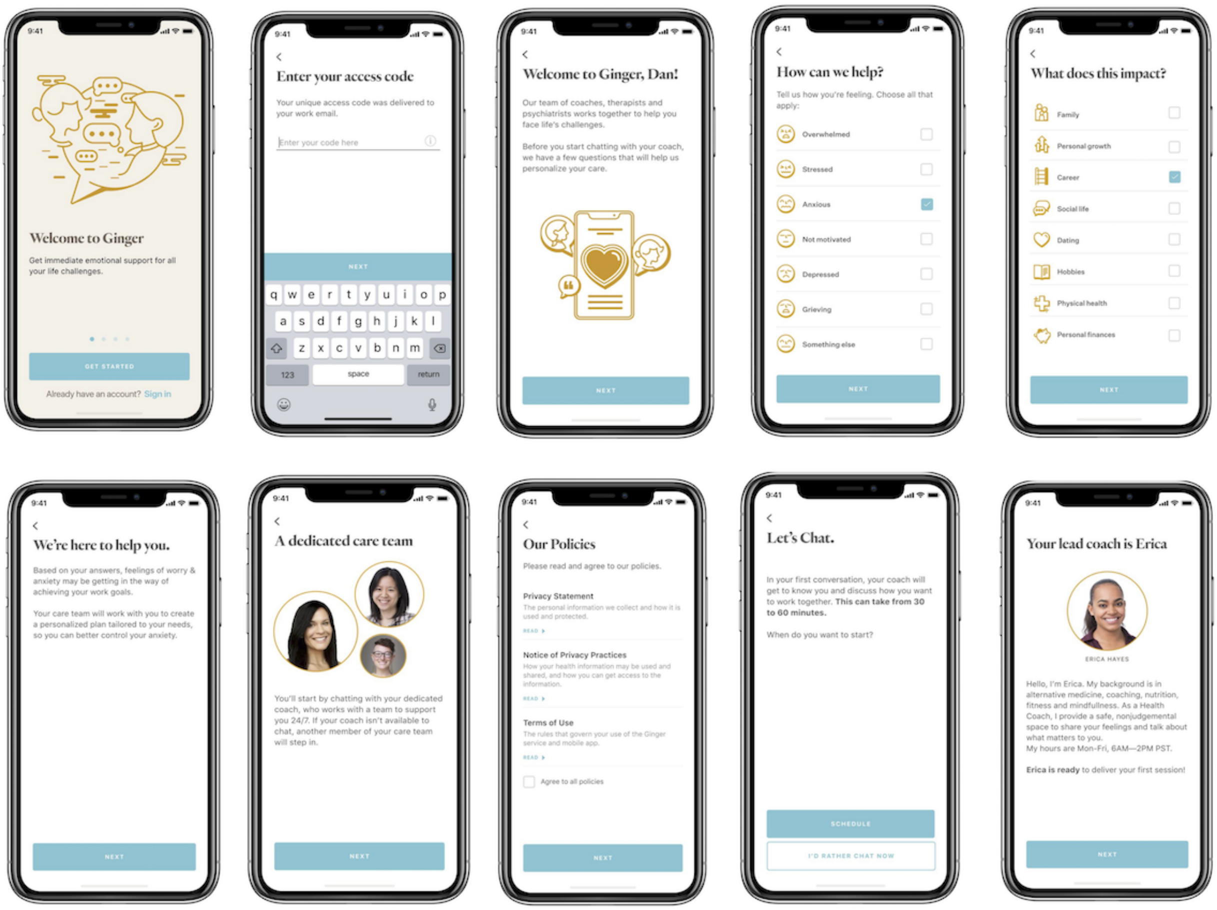
For this study, we included Ginger users aged 18 or older who downloaded the app between January 1, 2018, and December 31, 2019. This period was chosen as it reflects the approximate timing of when Ginger began to provide care to members via its employer business. Based on these criteria, 24,682 individuals were eligible for the study, 10,942 users (44.33%) completed the intake survey, and 1662 users (6.73%) completed a 12-week follow-up survey and were included in our analysis. Survey response rates are a challenge in both online studies and behavioral health practice; research has shown that the majority of participants in remote studies discontinue participation within the first week of a study and that online survey rates tend to be much lower than in-person [23,24]. The survey response rates here are lower than what is reported in the literature, which is expected given the observational nature of this study and the design of the survey system in a digital product, where users are not required to complete surveys to receive care.

Procedures

The Ginger System

Ginger provides members with behavioral health coaching, therapy, and psychiatry, along with self-guided content and assessments primarily via a mobile app platform. Members generally have access to Ginger as part of their employee or health plan benefits. After downloading the mobile app, users can start chatting with a behavioral health coach within minutes (Figures 1 and 2). Ginger coaches are full-time employees who typically have an advanced degree in a field related to mental health, and an accredited Coach Certification [25,26]. While many users solely engage with coaches, some will request or require escalation to psychotherapy services. Examples of situations that require escalation include individuals with chronic mental illness and severe trauma, the potential of harm to self or others, and significant mental instability (ie, hallucinations, delusions, and extreme mood swings). When members are escalated to therapy or psychiatry, they may continue working with a coach on an on-demand basis, provided that they also agree to seek additional care concurrently; coaching can continue supporting them in addressing day to day goals and challenges, as well as acting as an adjunct to the care plan put in place by their therapist or psychiatrist.

Figure 1. Ginger mobile app onboarding screens.



Welcome to Ginger

Get immediate emotional support for all your life challenges.

Already have an account? Sign in

Enter your access code

Your unique access code was delivered to your work email.

Welcome to Ginger, Dan!

Our team of coaches, therapists and psychiatrists works together to help you face life's challenges.

Before you start chatting with your coach, we have a few questions that will help us personalize your care.

How can we help?

Tell us how you're feeling. Choose all that apply:

- Overwhelmed ☐
- Stressed ☐
- Anxious ☑
- Not motivated ☐
- Depressed ☐
- Grieving ☐
- Something else ☐

What does this impact?

- Family ☐
- Personal growth ☐
- Career ☑
- Social life ☐
- Dating ☐
- Hobbies ☐
- Physical health ☐
- Personal finances ☐

We're here to help you.

Based on your answers, feelings of worry & anxiety may be getting in the way of achieving your work goals.

Your care team will work with you to create a personalized plan tailored to your needs, so you can better control your anxiety.

A dedicated care team

You'll start by chatting with your dedicated coach, who works with a team to support you 24/7. If your coach isn't available to chat, another member of your care team will step in.

Our Policies

Please read and agree to our policies.

Privacy Statement
The personal information we collect and how it is used and protected.

Notice of Privacy Practices
How your health information may be used and shared, and how you can get access to the information.

Terms of Use
The rules that govern your use of the Ginger service and mobile app.

☐ Agree to all policies

Let's Chat.

In your first conversation, your coach will get to know you and discuss how you want to work together. **This can take from 30 to 60 minutes.**

When do you want to start?

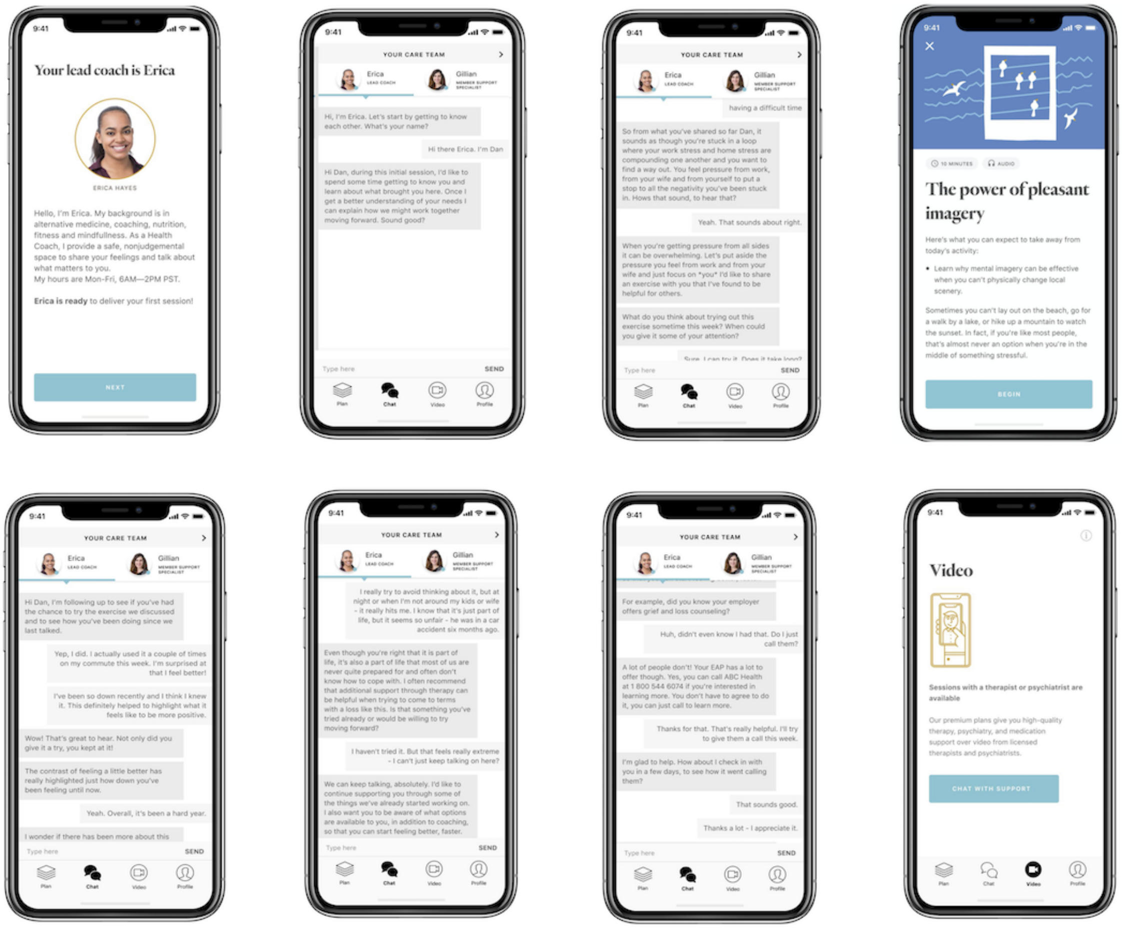
Your lead coach is Erica

ERICA HAYES

Hello, I'm Erica. My background is in alternative medicine, coaching, nutrition, fitness and mindfulness. As a Health Coach, I provide a safe, nonjudgemental space to share your feelings and talk about what matters to you.
My hours are Mon-Fri, 6AM—2PM PST.

Erica is ready to deliver your first session!

Figure 2. Ginger coach chat screens.



Your lead coach is Erica

ERICA HAYES

Hello, I'm Erica. My background is in alternative medicine, coaching, nutrition, fitness and mindfullness. As a Health Coach, I provide a safe, nonjudgemental space to share your feelings and talk about what matters to you.
My hours are Mon-Fri, 6AM—2PM PST.

Erica is ready to deliver your first session!

Hi, I'm Erica. Let's start by getting to know each other. What's your name?

Hi there Erica. I'm Dan

Hi Dan, during this initial session, I'd like to spend some time getting to know you and learn about what brought you here. Once I get a better understanding of your needs I can explain how we might work together moving forward. Sound good?

having a difficult time

So from what you've shared so far Dan, it sounds as though you're stuck in a loop where your work stress and home stress are compounding one another and you want to find a way out. You feel pressure from work, from your wife and from yourself to put a stop to all the negativity you've been stuck in. Hows that sound, to hear that?

Yeah. That sounds about right.

When you're getting pressure from all sides it can be overwhelming. Let's put aside the pressure you feel from work and from your wife and just focus on *you* I'd like to share an exercise with you that I've found to be helpful for others.

What do you think about trying out this exercise sometime this week? When could you give it some of your attention?

Sure, I can try it. Does it take long?

The power of pleasant imagery

Here's what you can expect to take away from today's activity:

- Learn why mental imagery can be effective when you can't physically change local scenery.

Sometimes you can't lay out on the beach, go for a walk by a lake, or hike up a mountain to watch the sunset. In fact, if you're like most people, that's almost never an option when you're in the middle of something stressful.

Hi Dan, I'm following up to see if you've had the chance to try the exercise we discussed and to see how you've been doing since we last talked.

Yep, I did. I actually used it a couple of times on my commute this week. I'm surprised at that I feel better!

I've been so down recently and I think I knew it. This definitely helped to highlight what it feels like to be more positive.

Wow! That's great to hear. Not only did you give it a try, you kept at it!

The contrast of feeling a little better has really highlighted just how down you've been feeling until now.

Yeah. Overall, it's been a hard year.

I wonder if there has been more about this

I really try to avoid thinking about it, but at night or when I'm not around my kids or wife - it really hits me. I know that it's just part of life, but it seems so unfair - he was in a car accident six months ago.

Even though you're right that it is part of life, it's also a part of life that most of us are never quite prepared for and often don't know how to cope with. I often recommend that additional support through therapy can be helpful when trying to come to terms with a loss like this. Is that something you've tried already or would be willing to try moving forward?

I haven't tried it. But that feels really extreme - I can't just keep talking on here?

We can keep talking, absolutely. I'd like to continue supporting you through some of the things we've already started working on. I also want you to be aware of what options are available to you, in addition to coaching, so that you can start feeling better, faster.

For example, did you know your employer offers grief and loss counseling?

Huh, didn't even know I had that. Do I just call them?

A lot of people don't! Your EAP has a lot to offer though. Yes, you can call ABC Health at 1 800 544 6074 if you're interested in learning more. You don't have to agree to do it, you can just call to learn more.

Thanks for that. That's really helpful. I'll try to give them a call this week.

I'm glad to help. How about I check in with you in a few days, to see how it went calling them?

That sounds good.

Thanks a lot - I appreciate it.

Video

Sessions with a therapist or psychiatrist are available

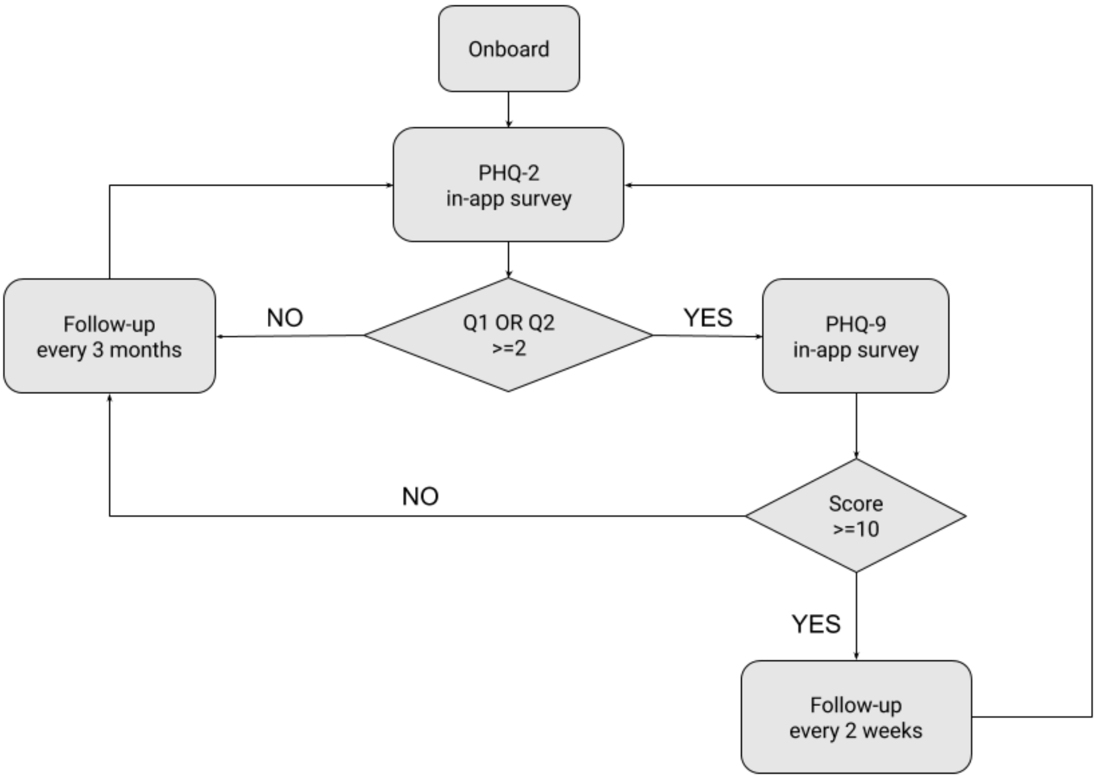
Our premium plans give you high-quality therapy, psychiatry, and medication support over video from licensed therapists and psychiatrists.

Data Collection

The Ginger system incorporates regular check-ins and feedback to understand better what solutions work for individual members, consistent with principles of measurement-based care (MBC). Although MBC has been demonstrated to enhance care, studies have shown that less than 20 percent of behavioral health practitioners integrate it into their practice [27]. Ginger assesses depression via the Patient Health Questionnaire (PHQ), delivered via the Ginger mobile app. The PHQ is one of the most validated assessments in mental health and is commonly used by clinicians in screening for and diagnosing depression in addition to monitoring treatment response [28]. When users onboard with the Ginger app (Figure 1), they are prompted first to answer the PHQ-2 questionnaire. Individuals that provide answers with a score of 2 or above for either question are considered a "positive depression screen" and then asked to complete the full PHQ-9.

There are no strict guidelines on how often to re-administer the PHQ-2 and PHQ-9; it can be re-administered as needed with a standard recommendation for monitoring and adjusting treatment every 4-6 weeks for users seeking treatment for depression [29]. Similarly, according to the US Preventive Services Task Force (USPSTF), there is little evidence regarding the optimal and interval for screening for depression. The USPSTF recommends a pragmatic approach of screening all adults who have not been screened previously and "using clinical judgment in consideration of risk factors, comorbid conditions, and life events to determine if additional screening of high-risk patients is warranted" [30,31]. Based on this guidance, Ginger administers the survey every 2 weeks for users with PHQ-9 scores ≥10 and every 3 months for users with PHQ-9 scores <10 in order to monitor symptom response and assess if additional care is warranted (Figure 3). Survey completion is not forced so as not to withhold support from members who require care.

Figure 3. Ginger PHQ survey logic.



In addition to clinical outcomes, Ginger collects and analyzes product data to understand a member's progress and journey through the system, including access, engagement, and member satisfaction.

Measures

Engagement

Engagement is characterized based on product user behavior data, including the number of coaching sessions and the number of clinical (therapy and psychiatry) appointments attended. A coaching session is defined as 25 messages exchanged. This threshold is based on an internal analysis of messaging patterns in addition to coaching team feedback on what constitutes a typical session. These measures provide insight into dropoff, adherence, and dosing, which are essential areas of study in this field [32,33].

Member Satisfaction

Member Satisfaction is measured using a 5-star survey. Ginger regularly prompts users to rate their experience via "star ratings." For coaching, this is required after the first 25 messages (ie, one text-based coaching session), and then every two weeks following. For clinical sessions, members are prompted to provide a rating after every therapy and psychiatry appointment. In addition to being a universal tool for consumers to rate products and services online (eg, Yelp, Amazon, Google), star ratings have also been used to understand provider care [34,35]. Although there are limitations associated with these consumer ratings, they can provide valuable insights for further exploration. Given the potential biases in these responses, the Ginger system includes proactive outreach from the member support team anytime there is less than a 4-star rating in addition to a clinical QA program to understand the care a member is receiving.

Depression Symptoms

Depression symptoms were assessed using the 2-item Patient Health Questionnaire (PHQ-2). The PHQ-2 comprises the first two items of the PHQ-9, assessing the degree to which an individual has experienced depressed mood and anhedonia over the preceding two weeks. A description of this survey logic is provided above. For this analysis, a negative depression screen means a user's response to each PHQ-2 question was less than 2 (ie, a response of "not at all" or "several days"). A positive depression screen means a user's response to either PHQ-2 question was ≥2 (ie, a response of "more than half the days" or "nearly every day"). Thus, a positive screen in this system can be interpreted to an individual having "little interest or pleasure in doing things" and "feeling down, depressed, or hopeless" more than half the days in the last 2 weeks. Although the PHQ-2 is generally used as a screening survey, it has a relatively high sensitivity for major depressive disorder, and previous research has utilized it as a longitudinal outcome measure [36-38]. Additionally, it provides insight into the severity of two key symptoms of depression: depressed mood and anhedonia. For this analysis, we considered a follow-up window of 8-12 weeks to reflect coaching and clinical protocols and the period in which we expect coaching and psychotherapy to have a measurable impact [32,33].

Data Management and Analysis

Data for this study were processed using Looker, business intelligence, and data analytics platform. Data were analyzed in Python and exported to spreadsheets for final analysis and review. We first assessed descriptive statistics for eligible users and those who completed follow-up. We then looked at the percentage of users demonstrating a change in PHQ-2 screen from baseline to follow-up for those who completed in-app surveys. We further evaluated these changes from baseline to follow-up using the McNemar test. The McNemar test compares dependent samples in terms of a dichotomous variable and is used for pretest/posttest designs or in time series data where the same sample is tested at least in two points in time [39]. We determined this to be the most appropriate test given the study design and nature of the outcome variable (comparing PHQ-2 screen outcomes at baseline and follow-up).

To evaluate outcomes by care modality (coaching only, therapy + psychiatry, hybrid) for users who screened positive at intake, we performed a chi-squared test to see if the proportion of individuals with reduced depression symptoms (change from a positive screen at intake to negative at follow-up) differed between these groups.

Ethics Statement

This study represents a secondary analysis of pre-existing de-identified data. The study team does not have access to the participants or participant identifying information and does not intend to recontact participants. This study protocol was reviewed by Advarra IRB and determined to be exempt from IRB oversight as de-identified secondary data analysis is generally not regarded as human subjects research.

Results

Table 1 reports descriptive characteristics of the sample who completed follow-up. Of 1662 individuals included in this analysis, 46.5% (772/1662) screened positive for depression at intake, and 53.6% (890/1662) screened negative according to the PHQ-2 questionnaire.

Gender and demographic data were missing for a large portion of the sample, as this has historically been optional information provided in employer eligibility files. For those users who did have data reported (N=678), 68.7% (466/678) were female, and 30.8% (209/678) were male. Of the individuals that reported age information, 11.5% (122/1064) were 18-24, 51.3% (546/1064) were 25-34, 22.7% (242/1064) were 35-44, 14.3% (152/1064) were 45-64, and 0.1% (2/1064) were 65 or older. When looking at the modality of care in which individuals were engaged, 65.5% (1089/1662) were only engaged in coaching, 27.6% (458/1662) were engaged in both coaching and clinical (therapy or psychiatry) services, 3.3% (54/1662) were only engaged in clinical services, and 3.7% (61/1662) used the app but did not engage in coaching or clinical services.

For those individuals engaged in coaching, the average number of sessions was 6 over the study period. For those individuals engaged in any clinical service (therapy or psychiatry), the average number of appointments was 6 over the period. The average number of therapy appointments for those engaged in therapy services was 6. The average number of psychiatry appointments for those engaged in psychiatry services was 3.

Table 2 shows the frequency and percentage of users by PHQ-2 result at intake and follow-up. Of the 1662 individuals who completed follow-up, 772 (46.5%) were considered a positive screen at intake, and 890 (53.6%) were considered a negative screen at intake. At the follow-up window, 477 (28.7%) of individuals screened positive, and 1185 (71.3%) screened negative. Using the McNemar test, we concluded that there was a statistically significant difference in the proportion of users with a positive PHQ-2 result at baseline and follow-up, $P<.001$.

Table 1. Characteristics of the study cohort.

Characteristic	Value
Gender, n (%)	
Female	466 (28.04)
Male	209 (12.58)
Other	3 (0.18)
No response	984 (59.21)
Age, n (%)	
18-24 years old	122 (7.34)
25-34 years old	546 (32.85)
35-44 years old	242 (14.56)
45-64 years old	152 (9.15)
65 or older	2 (0.12)
No response	598 (35.98)
Care modality, n (%)	
Coaching only	1089 (65.52)
Hybrid (coaching + clinical)	458 (27.56)
Clinical only	54 (3.25)
App only	61 (3.67)
Engagement, mean (SD)	
Coaching sessions	6.14 (5.44)
Clinical appointments	5.79 (3.12)
Therapy appointments	5.54 (2.8)
Psychiatry appointments	2.71 (1.7)
Member satisfaction, mean (SD)	
Coach star rating	4.63 (0.6)
Clinical star rating	4.74 (0.61)
Patient Health Questionnaire intake screen, n (%)	
Positive	772 (46.45)
Negative	890 (53.55)

Table 2. Patient Health Questionnaire-2 (PHQ-2) results at intake and follow-up (N=1662).

PHQ-2 screen result (pre)	Post		Count, n	X^2 (*df*)	*P* value
	Positive, n (%)	Negative, n (%)			
Positive	312 (18.77)	460 (27.68)	1662	139.24 (1)	<.001
Negative	165 (9.93)	725 (43.62)			

[a] McNemar chi-squared test.

Table 3 shows the frequency and percentage of users who experienced a change in their PHQ-2 result by care modality for the users who screened positive at intake (N=748). Using a chi-square test, we concluded that there was not a significant association between the proportion of individuals with a positive screen at intake who improved at follow-up and care modality (*P*=.77).

Table 3. Association between PHQ-2 improvements and care modality (N=748).

Variable (care modality)	Change, n (%)	No change, n (%)	Count, n	X^2 (*df*)	*P* value[a]
Coaching only	281 (58.54)	199 (41.46)	480	1.1289 (2)	.770
Therapy only	16 (61.54)	10 (38.46)	26		
Hybrid	150 (61.98)	92 (38.02)	242		

[a] Chi-square test for independence.

Discussion

Principal Findings

Although there is growing evidence that digital and virtual mental health interventions show promise in improving clinical outcomes for individuals with depression, many studies have highlighted the need to study further the features of specific technologies and the populations that use them. This study adds to this literature by describing a specific virtual on-demand mental health system and providing some preliminary results on user characteristics, engagement, and depression symptoms to guide future research.

Descriptive statistics showed that a majority of individuals were only utilizing coaching services (65.52%), and a minority of individuals were only utilizing clinical services (3.25%). The average coach and clinician star rating was >4.5, which provides some signal that members receiving care are satisfied with the care they are receiving, although more research is required to understand satisfaction with specific aspects of the system.

A McNemar test showed that the proportion of individuals with a positive PHQ-2 screen significantly decreased at follow-up, thus providing preliminary evidence that the Ginger system has an impact on decreasing depression symptom severity. A chi-square test concluded that there was not a significant association between care modality and the proportion of individuals with a positive screen improving at follow-up. While further research is required, this suggests that text-based behavioral health coaching alone can be an effective modality for reducing depression symptoms in addition to traditional clinical services like therapy and psychiatry. Based on the operationalization of our outcome variable, it is important to note that this analysis only evaluated change in PHQ-2 screen results; thus, we are underestimating the proportion of individuals who improved because we are not capturing individuals who screened positive at follow-up according to the PHQ-2 but still experienced a full symptom response according to PHQ-9.

Limitations

This analysis has various limitations based on historical product and clinical design of the system that generated these observational data. First, PHQ survey completion is not required for all users. Thus, the cohort we were able to study was limited, and these outcomes are not necessarily generalizable to the overall user base due to bias in the users who complete the survey. Efforts to improve the product design, user experience, and integrations with clinical workflows will also allow us to study a broader cohort in the future and make more generalizable conclusions.

Given the historical survey design of this system, clinical outcomes were operationalized as a binary variable using PHQ-2, which can be interpreted as a positive or negative depression screen or severity of anhedonia and depressed mood. Thus, we could only measure outcomes as a dichotomous variable and did not assess the extent of symptom response on a continuous scale. Although the PHQ-2 is generally intended to be a screening survey, it has a relatively high sensitivity for major depressive disorder, and there is precedent for using it as a longitudinal outcome [36-38]. Furthermore, it explicitly measures the severity of two key symptoms of depression: anhedonia and depressed mood.

Although we have some demographic data on users via employer eligibility files, missing data meant we were not able to stratify our analyses or control for specific demographic variables since users are not currently required to provide this information directly to Ginger. Future product updates will address this missing data issue and allow stratified analyses of outcomes by demographics (gender, age, socioeconomic status) to help us better understand how outcomes differ for specific populations.

Finally, because we lacked a control group, we are unable to understand the significance of these outcomes versus usual care or no care. Given ethical challenges, few prospective non-intervention studies following the natural course of untreated depression exist [37]. Prior research looking at wait-list control groups has suggested that 20-25 percent of untreated depression cases will remit within 3 months [40,41].

Collectively, these limitations point to many directions for future research. By using this dataset, follow-up studies could examine which aspects of this system are associated with changes in clinical outcomes, eg, different types of coaching and therapy, different thresholds of engagement, frequency of coach and clinical interaction, and different types of in-app content. New measurements and survey tools such as productivity, quality of life, and functional outcomes will also help evaluate impact in a broader cohort of users where the PHQ survey is not the most appropriate measure of progress, eg, those who screen negatively for depression. Finally, randomized controlled studies could also build upon this research and further test hypotheses on the efficacy of specific interventions in this system.

Conclusion

There is growing evidence that telehealth and other digital interventions can be useful in reducing symptoms of depression and other mental health conditions. This study adds to the literature by describing a specific on-demand mental health system and investigating utilization patterns and impact. The results of this exploratory study show a significant decrease in the proportion of users experiencing depressed mood and anhedonia at follow-up. Limitations with this study design mean that these results are not generalizable to the entire user base nor attributable to a specific intervention. Future studies can address these limitations and provide additional insight into which features of the system are most associated with outcomes in different populations.

Acknowledgments

SK authored this manuscript. MY conducted the data analysis. JH and WX provided clinical and technical expertise and supported manuscript review and edits.

Conflicts of Interest

All authors are current employees of Ginger.

References

1. National Institute of Mental Health. Major Depression URL: https://www.nimh.nih.gov/health/statistics/major-depression.shtml [accessed 2020-05-20]
2. Brody D, Pratt L, Hughes J. National Center for Health Statistics.: Centers for Disease Control and Prevention, National Center for Health Statistics Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013–2016 URL: https://www.cdc.gov/nchs/data/databriefs/db303.pdf [accessed 2020-06-04]
3. Blue Cross Blue Shield (BCBS). BCBS Health Index URL: https://www.bcbs.com/the-health-of-america/health-index [accessed 2020-05-04]
4. Chow W, Doane M, Sheehan J, Alphs L, Le H. Economic Burden Among Patients With Major Depressive Disorder: An Analysis of Healthcare Resource Use, Work Productivity, and Direct and Indirect Costs by Depression Severity. American Journal of Managed Care 2019:1-3 [FREE Full text]
5. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009;28(3):w490-w501. [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
6. Kaiser Family Foundation. Mental Health Care Health Professional Shortage Areas (HPSAs) URL: https://www.kff.org/other/state-indicator/mental-health-care-health-professional-shortage-areas-hpsas/ [accessed 2020-06-04]
7. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. Psychiatr Serv 2009 Oct;60(10):1323-1328. [doi: 10.1176/ps.2009.60.10.1323] [Medline: 19797371]
8. Waitzfelder B, Stewart C, Coleman K, Rossom R, Ahmedani B, Beck A, et al. Treatment Initiation for New Episodes of Depression in Primary Care Settings. J Gen Intern Med 2018 Aug;33(8):1283-1291 [FREE Full text] [doi: 10.1007/s11606-017-4297-2] [Medline: 29423624]
9. Lin P, Campbell DG, Chaney EF, Liu C, Heagerty P, Felker BL, et al. The influence of patient preference on depression treatment in primary care. Ann Behav Med 2005 Oct;30(2):164-173. [doi: 10.1207/s15324796abm3002_9] [Medline: 16173913]
10. Raue P, Schulberg H, Heo M, Klimstra S, Bruce M. Patients' depression treatment preferences and initiation, adherence, and outcome: a randomized primary care study. Psychiatr Serv 2009 Mar 01;60(3):337-343 [FREE Full text] [doi: 10.1176/appi.ps.60.3.337] [Medline: 19252046]
11. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Arch Intern Med 2006 Nov 27;166(21):2314-2321. [doi: 10.1001/archinte.166.21.2314] [Medline: 17130383]
12. Overbeck G, Davidsen AS, Kousgaard MB. Enablers and barriers to implementing collaborative care for anxiety and depression: a systematic qualitative review. Implement Sci 2016 Dec 28;11(1):165 [FREE Full text] [doi: 10.1186/s13012-016-0519-y] [Medline: 28031028]
13. Kaiser Family Foundation. 2018. One-Fourth Of Adults And Nearly Half of Adults Under 30 Don’t Have A Primary Care Doctor URL: https://www.kff.org/other/slide/one-fourth-of-adults-and-nearly-half-of-adults-under-30-dont-have-a-primary-care-doctor/ [accessed 2020-06-04]
14. Godleski L, Darkins A, Peters J. Outcomes of 98,609 U.S. Department of Veterans Affairs patients enrolled in telemental health services, 2006-2010. Psychiatr Serv 2012 Apr;63(4):383-385. [doi: 10.1176/appi.ps.201100206] [Medline: 22476305]
15. Coughtrey AE, Pistrang N. The effectiveness of telephone-delivered psychological therapies for depression and anxiety: A systematic review. J Telemed Telecare 2016 Dec 30;24(2):65-74. [doi: 10.1177/1357633x16686547]
16. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
17. Safavi K, Cohen A. Health Affairs. 2019 Feb 27. The Oversell And Undersell Of Digital Health URL: https://www.healthaffairs.org/do/10.1377/hblog20190226.63748/full/ [accessed 2020-06-04]

18. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/s2589-7500(19)30013-5]
19. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
20. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. Br J Sports Med 2014 Apr;48(8):731-736. [doi: 10.1136/bmj.f6753] [Medline: 24659611]
21. Marcelle ET, Nolting L, Hinshaw SP, Aguilera A. Effectiveness of a Multimodal Digital Psychotherapy Platform for Adult Depression: A Naturalistic Feasibility Study. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e10948 [FREE Full text] [doi: 10.2196/10948] [Medline: 30674448]
22. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the Wild: Using Novel Data to Examine the Reach, Use, and Impact of PTSD Coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
23. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? Assessment & Evaluation in Higher Education 2008 Jun;33(3):301-314. [doi: 10.1080/02602930701293231]
24. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
25. National Board for Health and Wellness Coaching. URL: https://nbhwc.org/ [accessed 2020-06-04]
26. Institute of Coaching. Introduction to the Coaching Profession URL: https://www.instituteofcoaching.org/resources/introduction-coaching-profession [accessed 2020-06-04]
27. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing Measurement-Based Care in Behavioral Health: A Review. JAMA Psychiatry 2019 Mar 01;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
28. American Psychological Association. Patient Health Questionnaire (PHQ-9 & PHQ-2) URL: https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/patient-health [accessed 2020-06-04]
29. Simon GE, Rutter CM, Peterson D, Oliver M, Whiteside U, Operskalski B, et al. Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? Psychiatr Serv 2013 Dec 01;64(12):1195-1202 [FREE Full text] [doi: 10.1176/appi.ps.201200587] [Medline: 24036589]
30. Maurer DM, Raymond TJ, Davis BN. Depression: Screening and Diagnosis. Am Fam Physician 2018 Oct 15;98(8):508-515 [FREE Full text] [Medline: 30277728]
31. United States Preventive Services Taskforce. Final Recommendation Statement: Depression in Adults: Screening URL: https://www.uspreventiveservicestaskforce.org/uspstf/document/RecommendationStatementFinal/depression-in-adults-screening [accessed 2020-06-04]
32. Nieuwsma JA, Trivedi RB, McDuffie J, Kronish I, Benjamin D, Williams JW. Brief Psychotherapy for Depression: A Systematic Review and Meta-Analysis. Int J Psychiatry Med 2012 Jun 15;43(2):129-151. [doi: 10.2190/pm.43.2.c]
33. Maeng D, Cornell AE, Nasra GS. Utilization of an Employee Behavioral Health Program and Its Effects on Outcomes for Depression and Anxiety Disorders. Journal of Occupational and Environmental Medicine 2019;61(10):812-817. [doi: 10.1097/jom.0000000000001678]
34. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care. Health Aff (Millwood) 2016 Apr;35(4):697-705 [FREE Full text] [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]
35. Donnally CJ, Roth ES, Li DJ, Maguire JA, McCormick JR, Barker GP, et al. Analysis of Internet Review Site Comments for Spine Surgeons. SPINE 2018;43(24):1725-1730. [doi: 10.1097/brs.0000000000002740]
36. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
37. Fleishman J, Zuvekas S, Pincus H. Agency for Healthcare Research and Quality. 2014 Oct. Screening for Depression Using the PHQ-2: Changes over Time in Conjunction with Mental Health Treatment URL: https://meps.ahrq.gov/data_files/publications/workingpapers/wp_14002.pdf [accessed 2020-06-04]
38. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
39. Fisher MJ, Marshall AP, Mitchell M. Testing differences in proportions. Aust Crit Care 2011 May;24(2):133-138. [doi: 10.1016/j.aucc.2011.01.005] [Medline: 21536451]
40. Casacalenda N, Perry JC, Looper K. Remission in major depressive disorder: a comparison of pharmacotherapy, psychotherapy, and control conditions. Am J Psychiatry 2002 Aug;159(8):1354-1360. [doi: 10.1176/appi.ajp.159.8.1354] [Medline: 12153828]
41. Whiteford HA, Harris MG, McKeon G, Baxter A, Pennell C, Barendregt JJ, et al. Estimating remission from untreated major depression: a systematic review and meta-analysis. Psychol. Med 2012 Aug 10;43(8):1569-1585. [doi: 10.1017/s0033291712001717]

Abbreviations

CMS: Centers for Medicare & Medicaid Services
MBC: measurement-based care
PHQ: Patient Health Questionnaire

Edited by G Eysenbach, J Torous; submitted 13.02.20; peer-reviewed by MW Haun, J Hensel, T Campellone, CY Lin; comments to author 10.03.20; revised version received 04.05.20; accepted 20.05.20; published 18.06.20.

Please cite as:
Kunkle S, Yip M, Ξ W, Hunt J
Evaluation of an On-Demand Mental Health System for Depression Symptoms: Retrospective Observational Study
J Med Internet Res 2020;22(6):e17902
URL: http://www.jmir.org/2020/6/e17902/
doi:10.2196/17902
PMID:32554387

Original Paper

A Blended Physiotherapy Intervention for Persons With Hemophilic Arthropathy: Development Study

Merel A Timmer[1], PT, PhD; Corelien J J Kloek[2], PT, PhD; Piet de Kleijn[1], PT; Isolde A R Kuijlaars[1], MSc, PT; Roger E G Schutgens[1], MD, PhD; Cindy Veenhof[2,3], PT, PhD; Martijn F Pisters[3], PT, PhD

[1]van Creveldkliniek, University Medical Center Utrecht, Utrecht, Netherlands

[2]Research Group Innovation of Human Movement Care, University of Applied Sciences Utrecht, Utrecht, Netherlands

[3]Physical Therapy Research, Department of Rehabilitation, Physical Therapy Science and Sport, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands

Corresponding Author:
Merel A Timmer, PT, PhD
van Creveldkliniek
University Medical Center Utrecht
Heidelberglaan 100
Utrecht, 3508 GA
Netherlands
Phone: 31 88 755 1709
Email: m.a.timmer@umcutrecht.nl

Abstract

Background: Joint bleeds are the hallmark of hemophilia, leading to a painful arthritic condition called as hemophilic arthropathy (HA). Exercise programs are frequently used to improve the physical functioning in persons with HA. As hemophilia is a rare disease, there are not many physiotherapists who are experienced in the field of hemophilia, and regular physiotherapy sessions with an experienced physiotherapist in the field of hemophilia are not feasible for persons with HA. Blended care is an innovative intervention that can support persons with HA at home to perform the advised physical activities and exercises and provide self-management information.

Objective: The aim of this study was to develop a blended physiotherapy intervention for persons with HA.

Methods: The blended physiotherapy intervention, namely, e-Exercise HA was developed by cocreation with physiotherapists, persons with HA, software developers, and researchers. The content of e-Exercise HA was compiled using the first 3 steps of the Center for eHealth Research roadmap model (ie, contextual inquiry, value specification, and design), including people with experience in the development of previous blended physiotherapy interventions, a literature search, and focus groups.

Results: A 12-week blended intervention was developed, integrating face-to-face physiotherapy sessions with a web-based app. The intervention consists of information modules for persons with HA and information modules for physiotherapists, a graded activity program using a self-chosen activity, and personalized video-supported exercises. The information modules consist of text blocks, videos, and reflective questions. The patients can receive pop-ups as reminders and give feedback on the performance of the prescribed activities.

Conclusions: In this study, we developed a blended physiotherapy intervention for persons with HA, which consists of information modules, a graded activity program, and personalized video-supported exercises.

(J Med Internet Res 2020;22(6):e16631) doi:10.2196/16631

KEYWORDS

hemophilia; physiotherapy; exercise; eHealth; blended care; mobile phone

Introduction

Hemophilia is an X-linked congenital disorder that impairs the body's ability to make blood clots [1]. It is a rare disease, with a prevalence of 1 in 10,000 persons. Hemophilia is characterized by prolonged bleeding after injuries, easy bruising, and an increased risk of bleeding in the joints and muscles. When untreated, persons with severe hemophilia (<1% blood clotting factors) experience spontaneous bleeding, whereas persons with mild hemophilia (>5% blood clotting factors) experience bleeding only after a trauma [1]. Despite the introduction of prophylactic clotting factor replacement therapy in developed countries in the 1970s, approximately 2 joint bleeds per year are still observed in persons with severe hemophilia receiving prophylactic treatment [2]. The most affected body parts in persons with hemophilia are the ankle joints (42%), knees (20%), and elbows (20%) [3].

Recurrent bleeding in the joints eventually causes hemophilic arthropathy (HA). The pathogenesis of HA has certain features in common with inflammatory joint diseases, but HA is predominantly a degenerative joint disease, which is comparable to osteoarthritis (OA) [4]. The clinical symptoms of pain and the limited range of motion of the joints and atrophy of the surrounding muscles in HA are similar to those reported in OA. Furthermore, people with HA and OA experience limitations in their activities and participation and they are less active than their healthy peers [5].

A wide variety of exercise programs have been developed for persons with hemophilia [6]. However, since these interventions have shown mixed success, a preferred exercise intervention for persons with HA is still undecided. Moreover, regular physiotherapy is not feasible for many persons with hemophilia because physiotherapy treatment is often not covered by health insurance and physiotherapists experienced in hemophilia care are scarce. Since the symptoms and pathogenesis of HA are similar to those of OA, it can be hypothesized that increasing physical activity in daily life has the potential to improve the physical functioning in persons with HA, as previously shown in persons with OA [7]. To our knowledge, no interventions that target the daily movement behavior of persons with hemophilia have been described to date.

Medical apps in smartphones are an upcoming phenomenon in health care and they offer unique possibilities for providing information, supporting behavioral change, and encouraging self-control. Integrating a medical app in smartphones with regular face-to-face physiotherapy sessions can support the participants in performing the advised physical activity behavior and prescribed home exercises and support self-management beyond the walls of the physiotherapy practice. Previously developed electronic health (eHealth) apps for persons with hemophilia include videoconferencing apps and apps that report bleeding events and the use of clotting factor replacement therapy [8,9]. An eHealth intervention directed at changing the movement behavior and improving physical functioning is not yet available for persons with hemophilia.

Recently, a blended exercise program “e-Exercise” was developed for persons with OA, integrating face-to-face physiotherapy with a smartphone app [10]. This program consists of information modules on increasing self-management, personalized exercises, and a behavioral graded activity approach. A clustered randomized controlled trial showed that the effectiveness of e-Exercise in patients with OA was comparable to that of the usual physiotherapy in these patients [11]. Moreover, patients with OA in the e-Exercise group needed only 5 face-to-face sessions, whereas patients in the usual physiotherapy group needed 12 face-to-face sessions.

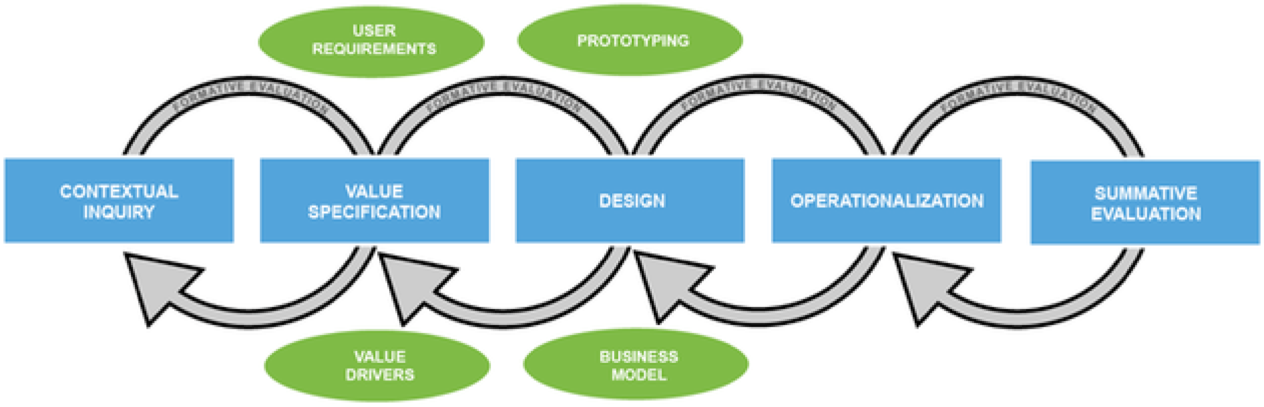
A blended physiotherapy intervention is expected to be beneficial for persons with HA because it has the potential to support behavioral change and it creates the opportunity to treat patients who have limited access to specialized physiotherapy. Furthermore, it could be used to support physiotherapists in primary health care, who are less familiar and not experienced in treating persons with HA. We hypothesized that certain aspects of the previously developed e-Exercise OA intervention could be suitable for persons with HA as well, given the previously mentioned similarities between the two disorders. However, significant changes might be needed in the intervention for HA because of the differences in the congenital character of hemophilia, possible fear of bleeding, involvement of multiple joints, and involvement of other joints (eg, ankles). The aim of this study was to develop a blended physiotherapy intervention for persons with HA. The proposed program was developed upon the existing e-Exercise intervention for persons with OA. During the development of this program, the similarities and differences in the treatment of persons with OA and HA were investigated.

Methods

Theoretical Model Roadmap

The theoretical Center for eHealth Research (CeHRes) roadmap model was used to develop e-Exercise HA [12]. The CeHRes roadmap is designed for the development and research of eHealth technologies. Cocreation plays a central role in this approach because it anticipates the needs and values of the stakeholders to improve the uptake and the effect of the intervention. Figure 1 shows the 5 stages of the CeHRes roadmap, which consists of connecting formative evaluation cycles. This study focused on the following first 3 stages of the model: contextual inquiry, value specification, and design.

Figure 1. CeHRes roadmap. CeHRes: Center for eHealth Research.



Stages 1 and 2: Contextual Inquiry and Value Specification

The aim of contextual inquiry and value specification is to establish the most important needs, values, and requirements of the stakeholders in relation to a blended physiotherapy intervention for persons with HA. The input in this step was based on (1) a literature search, (2) experience with the development and implementation of the e-Exercise OA intervention, and (3) focus groups with persons with HA and physiotherapists experienced in the field of hemophilia. These 3 points have been described in detail as follows.

First, a literature search was performed for studies published in PubMed till April 2019 using the search terms “hemophilia” AND “exercise” OR “physiotherapy.” Additionally, we searched for national and international physiotherapy guidelines on hemophilia and OA through patient societies and professional associations and by using search terms “physiotherapy” AND “guidelines and hemophilia” OR “osteoarthritis.”

Second, experience with the development and implementation of the e-Exercise OA intervention was established from previous studies [10,13,14] and by involving the authors of these previous studies in the development process.

Third, 2 focus groups were formed: one group consisted of 5 persons with HA and the other group consisted of 5 physiotherapists experienced in the field of hemophilia. Sessions with the focus groups lasted for around 1 hour and were moderated by MAT. A topic guide was used to discuss the following subjects: the preferred content of physiotherapy for persons with HA, advantages and barriers of usual physiotherapy, information provided and needed, thoughts and beliefs around physical activity (only persons with hemophilia), possible subgroups of persons with HA (only physiotherapists), and advantages and barriers to the use of eHealth. All focus groups were audio recorded and transcribed verbatim. The results were analyzed using a thematic analysis. The data analysis involved reading the transcripts, mapping quotes into codes with the open coding process, and summarizing the codes into themes iteratively.

Thus, information acquired by literature research, experience with a previously developed blended intervention for patients with OA, and focus groups were used to compile the content of the intervention. The authors compiled the first draft of the program. Consequently, all the members of the focus groups individually provided feedback on the draft. Patients were invited to give feedback on the information modules for patients, while physiotherapists were invited to comment on the information modules for physiotherapists. In addition, a hematologist and a social worker commented on the information modules. The physiotherapists collected the feedback with their coworkers in primary care, while patients discussed with their partners or their significant other. The comments were processed by the authors and a second draft was presented to the stakeholders. This process continued until all the members were satisfied.

Stage 3: Design

During the design phase, the researchers cooperated with “The Health Train,” a commercial eHealth entrepreneur, which was already involved in the previous developed e-Exercise OA app. Currently, e-Exercise interventions are also being developed and studied for low back pain, medically unexplained physical symptoms, and complaints of the neck and shoulder. Knowledge about the design and the functionality of the app studied during the development of other e-Exercise interventions were used for the development of e-Exercise HA. The collaboration with a commercial eHealth entrepreneur can facilitate a long-term implementation of the e-Exercise interventions in physiotherapy care, irrespective of research funding. The possibility of using the same platform for interventions for different patient populations and connecting e-Exercise with the most widely used electronic medical record by physiotherapists in primary care will facilitate the use of e-Exercise by physiotherapists. The research team provided input for the content of the web-based app, and “The Health Train” integrated this content into their platform.

Results

Stages 1 and 2: Contextual Inquiry and Value Specification

A recent Cochrane systematic literature review in 2016 [6] provided information on the effectiveness and safety of several exercise protocols for persons with hemophilia, including resistance exercises, isometric exercises, bicycle ergometry, treadmill walking, and hydrotherapy. Most interventions improved one of the following outcomes: pain, range of motion, strength, or walking tolerance. Functional exercises seemed more effective than static exercises for improving strength. No adverse events were reported as a result of any of the interventions. An additional literature search between December 2016 and April 2019 with the search terms “hemophilia” AND “exercise” OR “physiotherapy” yielded 4 original papers that reported similar results after different forms of exercise therapy and manual therapy [15-19]. No guidelines on physiotherapy for persons with HA were available. Instead, parts of the general guidelines for the treatment of hemophilia and parts of the Dutch physiotherapy guidelines for OA were used [1,20]. The demographic characteristics of the participants in both the focus groups are shown in Table 1 and Table 2.

The focus groups provided information on the following themes: current treatment for HA, attitude toward physical activity with HA, information requirements for HA, and values and requirements of e-Exercise for HA.

Table 1. Focus Group 1: Demographic characteristics of the participating patients.

Participants with HA[a]	Age (years)	Severity	Hemophilia A/B	Hemophilia joint health score[b]	Hemophilia activity list[c]
Participant 1	56	Severe	A	38	59
Participant 2	54	Severe	A	12	72
Participant 3	67	Severe	A	50	40
Participant 4	40	Severe	A	31	94
Participant 5	76	Severe	A	n.a.[d]	n.a.

[a] HA: hemophilic arthropathy.

[b] High score indicates worse joint status.

[c] High score indicates less limitations in activities, maximum score =100.

[d] n.a.: not assessed.

Table 2. Focus Group 2: Demographic characteristics of the participating physiotherapists.

Participants	Overall experience (years)	Experience with hemophilia (years)	Number of persons with hemophilia (n) that were treated	Work setting	Specialization
Physiotherapist 1	30	>10	>10	Primary care	Manual therapy
Physiotherapist 2	26	>10	0-5	Primary care	General
Physiotherapist 3	18	>10	0-5	Primary care	General
Physiotherapist 4	38	>10	>10	Primary care	Manual therapy
Physiotherapist 5	40	>10	>10	Hemophilia treatment center	Manual therapy

Current Treatment for Hemophilic Arthropathy

According to the physiotherapists, the basis of the treatment for persons with HA is similar to that of the treatment for patients with OA, and the same physiological principle in OA treatment should be used in HA physiotherapy treatment. The focus group session with the physiotherapists directed at the similarities and the differences between the preferred treatments for OA and HA. This focus group revealed that the treatment for both OA and HA has changed over the last few years from a hands-on approach to a coaching-based therapy, thereby enabling patients to achieve more self-control. The most important difference between OA and HA is that the physiotherapists need to be aware of the symptoms that can indicate a bleeding event while treating a person with HA. Another difference is that persons with HA have more knowledge about their condition—often even more than their primary care physiotherapist. This knowledge plays an important role in their coaching roles. Furthermore, persons with OA have to be motivated to be more active while persons with HA need to learn how to increase the intensity and the duration of the activities and exercises slowly and to avoid peak loads. Patients stated that they benefitted from the exercises and advice on movement behavior as well as the manual therapy. Focus group sessions with persons with HA revealed that the most critical barriers to adhering to physiotherapy were the limited reimbursement by the health insurance, execution of boring exercises, and stubbornness of the patients to accept advise. The patients mentioned that a good relationship with their physiotherapist facilitated their adherence to the physiotherapy treatment.

Attitude Toward Physical Activity

Persons with HA explained that they continuously balanced the risk of increasing complaints with the benefits of performing an activity. Patients preferred to do meaningful activities such as walking or cycling in their daily life. They stated that the proud feeling after accomplishing an activity goal motivated them to continue to be active. Some patients mentioned that they preferred to do exercises as well. The important conditions for achieving an activity goal are the choice of the type of activity and tailoring the increase in the intensity and duration with professional guidance. Patients were motivated to perform an activity if they received encouragement but they were demotivated when their complaints were not empathized by others. Patients expressed different opinions about consuming painkillers. Some stated that painkillers enabled them to be active and to relax, while others were afraid that the consumption of painkillers would make them overuse the joint and consequently increase the joint damage.

Requirement of Specific Information on Hemophilic Arthropathy

Both persons with HA and physiotherapists in primary care emphasized the need for specific information for physiotherapists in primary care. Physiotherapists in primary care need to know the basic information about hemophilia along with the specific dos and don'ts. Patients did not specify the need for more information for themselves. The physiotherapists stated that basic information about hemophilia need not be provided for such patients. Instead, the patients could benefit from information on the course of their complaints, differences between joint bleeding and arthropathic complaints, how to actively manage HA, and how to cope with the limitations in their daily activities and participation. Short videos were considered as the preferable mode of conveying information to persons with HA and physiotherapists. Information for patients must be provided in a simple language. Physiotherapists were prepared to spend no more than 30 minutes to read or hear all the information that was required. The patients appreciated the communication about their specific situation between the hemophilia treatment center and the primary care physiotherapist.

Values and Requirements of Blended Physiotherapy for Persons With Hemophilic Arthropathy

Both physiotherapists and persons with HA believed that an e-Exercise app could add positively to the current physiotherapy treatment for persons with HA. The physiotherapists stated that the frequency of face-to-face visits in this program should be high initially and then decreased during the intervention. The overall number of face-to-face visits should be personalized, which can be determined by the self-management skills and the restraining factors of the patient. The patients stated that the e-Exercise app should include choices for exercises and a variability in the exercises. The patients preferred to set a goal, perform meaningful activities (walking or biking), give feedback on how well they were able to follow the program, and receive pop-ups and an exercise scheme to increase adherence. Physiotherapists stated that they could benefit from a feature in the app through which they could communicate with the patients and be able to consult the hemophilia treatment center for questions. The illustrative quotes for the different themes are presented in Table 3.

Table 3. Quotes from physiotherapists and persons with HA[a].

Theme	Illustrative quote
Current treatment for HA	*…We have our principles of physiology and training. In strength training, overload is needed to make gains for the muscle fibers. This is no different for a person with hemophilia; however, you need to be more careful with increasing intensity.* [Physiotherapist]
Attitude toward physical activity with HA	*…When you have hemophilia, physical activity needs to involve mainly natural movements and things that I just do myself. I try that, I ride my bike, I walk, and I do not take the elevator but the stairs.* [Patient]
Information requirements	*…The difference between patients with OA and those with HA is that because persons with HA have had hemophilia all their life, they have more knowledge about their disease than their 63-year-old neighbor with a painful knee."* [Physiotherapist]
Values and requirements of e-Exercise	*…So you have complaints and you cannot do certain things. ...the physiotherapist asks what your most important feasible goal is. "…" if you have a problem with your elbow and you cannot put on your sock, it may only be a minimal difference, that can be very inconvenient, and if it is your goal to be able to do that.* [Patient]

[a] HA: hemophilic arthropathy.

Stage 3: Design

The e-Exercise HA intervention was developed based on the existing e-Exercise OA intervention because of the findings generated from our focus groups, in which the physiotherapists stated that the basis of the treatment for persons with OA and HA is similar. The core components of e-Exercise OA are self-management information, a graded activity approach, and personalized exercises [11]. The need for information modules in e-Exercise HA intervention was expressed by the focus group with physiotherapists as the physiotherapists felt that they lacked the knowledge to educate persons with HA. Therefore, self-management information modules for patients and information modules for physiotherapists were included within e-Exercise HA. Furthermore, focus groups revealed the need for slowly increasing an activity and avoiding peak loads, goal-setting and performing meaningful daily activities. Those needs were met by the graded activity approach. Personalized exercises were suggested by persons with HA and have been reported in a previous study [6]. Feedback from the stakeholders on the drafts mainly consisted of changes to the content of the information modules and the script of the information videos. Since the web-based environment was already developed in cocreation with the end users and adjusted based on the experiences during the trial, the design of the web-based environment was not changed.

e-Exercise Hemophilic Arthropathy

This 12-week intervention integrates face-to-face physiotherapy with a smartphone app. Physiotherapists can log into the website, which offers the platform with the physiotherapist-specific information about HA. The physiotherapists can create an account for each patient. In this account, physiotherapists can create and personalize an exercise program for each patient and monitor the progress. The website allows physiotherapists to continuously adapt the treatment program according to the needs of the individual patient. Patients can log into their personal account in a smartphone app. In this app, the patients receive weekly modules with information and are able to review their personal physical activity program and their personalized exercises. They can give feedback on the performance of the prescribed program and receive tailored feedback messages. The components of e-Exercise HA have been further explained below. A screenshot of the e-Exercise HA app is shown in Multimedia Appendix 1.

Self-Management Information Modules

The information modules consist of text blocks supported by short videos [1,6,20-24]. For the physiotherapists, the following 5 information modules are available: (1) What is hemophilia? (2) Development of HA. (3) HA and physiotherapy. (4) Identifying a bleed. (5) Organization of care. The patients receive the following information modules on their smartphone app every week: (1) What is HA? (2) Pain and HA. (3) Joint bleed or HA? (4) Physical activity and HA. (5) Responsible exercise. (6) Arthropathy management. (7) A physical activity plan. (8) Maintaining an active life. (9) Maintaining healthy body weight. (10) Dealing with the surrounding people. (11) Fatigue. (12) What can you do? After every module, patients are invited to answer a reflective question in that particular theme. Physiotherapists are encouraged to further discuss the themes during the next face-to-face visit. When the patients feel that a specific theme does not suit their personal situation, they can choose to skip that specific information module.

Graded Activity

The behavioral graded activity approach is directed at increasing the amount of physical activity in a time-contingent way so that the activities are gradually increased by present quotas [25]. First, the persons with HA select their most problematic activity (eg, walking). The initial baseline measures, in which the patients perform the selected activity at home to the limit of tolerance (pain-contingent), are then performed for a week and will be registered in the app. Second, after these baseline measurements are recorded, the patients set their individual treatment goal for the selected activity. During this procedure of goal setting, the physiotherapist acts as a coach only, as it is essential that the goal is the patient's own internal goal. Throughout the program, this activity has to be performed 3 times a week for 11 weeks. The duration will be gradually increased in a time-contingent manner to reach the final goal at the end of the program. The weekly program will be automatically generated starting at 50% of the baseline assessment. On self-chosen moments (eg, every Monday at 8 PM), patients receive reminders on their smartphone to perform this activity. Patients can turn the timer on and off when they perform their activities so that the physiotherapist can monitor the progression. Third, tailored feedback messages are provided at the end of the activities with the aim of motivating or slowing down the patients with their program by using the principles of operant conditioning. The progression will be evaluated during the face-to-face visits.

Exercises

e-Exercise HA is linked to a database of over 1500 exercises, which are supported with texts and videos. In the face-to-face sessions, physiotherapists can start an exercise program by choosing personalized exercises and parameters. On their apps, patients can recall their exercises, including the parameters and the information about their performance. Similar to that in the graded activity approach, the physiotherapist is able to monitor the performance of the exercises. During the face-to-face visits, the performance of the exercises can be evaluated and adjusted when needed.

Face-to-Face Sessions

Face-to-face sessions are conducted to recall web-based self-management modules and to discuss how the themes affect the patient's personal situation by evaluating the answers on the reflective questions. The progress in the graded activity program and in the performance of the exercises can be visualized in the management portal of the website used by the physiotherapist. When needed, the graded activity program and the prescribed exercises can be adjusted. We propose 6 face-to-face visits during the 12-week program, with a decreasing frequency as the program progresses. The visits will be proposed to be scheduled on weeks 1, 2, 4, 6, 9, and 12. However, the number of visits and the timing of the visits depend on the personal needs of the patients, including the self-management skills and digital skills.

Discussion

Principal Findings

In this study, we developed a blended physiotherapy intervention for persons with HA, which meets the needs of the persons with HA and the physiotherapists in primary care. To enable successful implementation, the intervention was developed by cocreation with stakeholders such as persons with HA, physiotherapists, software developers, and researchers. Our focus group study with persons with HA and physiotherapists yielded the following findings: physiotherapy for persons with HA is similar to that for persons with OA, specific tailored information on rare diseases such as hemophilia and HA is required for patients and physiotherapists, patients prefer meaningful daily life activities and want to work toward achieving a goal, patients prefer choices and variability in exercises, and the exercise program needs to be personalized. Both physiotherapists and persons with HA agreed that integration of an e-Exercise app within physiotherapy treatment could be beneficial for the treatment of HA. The most important difference between providing treatment for OA and HA is that physiotherapists need to be aware of the disease symptoms that can be indicative of bleeding when treating a person with HA.

Comparison With Other Studies

Currently, several eHealth apps are being developed and used in hemophilia care [26]. However, these are mainly videoconferencing apps and apps the report bleeding and the use of clotting factor replacement therapy. No study has yet described smartphone apps that support physiotherapy in persons with hemophilia or related bleeding disorders [26]. Focus groups in this study showed that the Dutch physiotherapists and persons with HA showed a positive attitude toward the use of a blended physiotherapy intervention. This is in accordance with previous studies showing that blended eHealth intervention is feasible and effective in persons with OA, although some physiotherapists expected to face some barriers in this intervention mode [11,14,27]. An important barrier for physiotherapists to use e-Exercise OA was that it could be used for only 1 disorder. The current expanding of the platform with e-Exercise programs for several disorders is expected to facilitate the use of e-Exercise HA [28]. Another barrier was that some physiotherapists felt that the program interfered with their professional autonomy. However, given the limited knowledge of primary care physiotherapists about hemophilia, it is expected that this is less of a problem for e-Exercise HA. In accordance with previous studies, the requirement of e-Exercise determined in this study included a smartphone app with information for patients, goal-setting features, and a personalized exercise program [10]. The need for information modules for physiotherapists is specific for hemophilia, which is understandable, given the rare nature of the disease.

Strengths and Limitations of This Study

Hampered uptake and implementation of eHealth solutions in clinical practice is a common problem [29]. A strength of this study is that the intervention was developed in cocreation with the end users (ie, patients and physiotherapists) according to the CeHRes roadmap in order to accelerate implementation at a later stage. Adding information modules particularly for physiotherapists was a strong recommendation by the stakeholders and is expected to improve the quality of care. A limitation of this study is that only Dutch stakeholders were invited. We expect that different barriers and facilitators might be present in different countries, thereby limiting the generalizability of our findings in different countries. Moreover, only patients aged over 40 years participated in the process of cocreation. This limits the generalizability of the findings to adolescents and young adults. Furthermore, the feasibility and the effectiveness of this intervention have not yet been investigated and will be investigated at a later stage of this project. The intervention in this study was built upon an existing intervention for persons with OA and is part of a platform for blended physiotherapy interventions for several disorders. This has enabled us to learn from the results of these interventions and will lower the threshold for the implementation of our intervention for persons with a rare disease. The use of an existing platform also limited us to make major changes to the web-based environment. This intervention has been developed in cocreation with stakeholders as well and we expect that preferences regarding this topic will not be very different from those of our population. The cocreation process in this study focused mainly on the content of the intervention and less on the web-based interface.

Clinical Implications and Future Research

A possible use of e-Exercise HA is by physiotherapists who are associated with a hemophilia treatment center and who are able to treat patients who visit the clinic a few times, but not on a weekly basis. Moreover, e-Exercise HA can support physiotherapists in primary care, who have no experience with hemophilia but can treat persons with HA locally. We expect blended physiotherapy to lead to less face-to-face visits. This could increase the accessibility of physiotherapy for persons with HA, leading to lower costs and more flexibility for patients. Furthermore, we expect that e-Exercise HA will improve self-management with respect to the movement behavior because patients are provided with information modules and are stimulated to perform activities at home. Future research should be directed toward investigating the effectiveness of this intervention and the implementation of this intervention in the Netherlands. To enable the international implementation of this program, the barriers and the facilitators of the stakeholders in different countries need to be investigated.

Conclusion

We developed a 12-week blended physiotherapy intervention that integrates face-to-face physiotherapy sessions with a smartphone app for persons with HA. This personalized intervention consists of self-management information, a graded self-chosen activity program, and video-supported personalized exercises, which integrates face-to-face sessions and a smartphone app.

Acknowledgments

This study was supported by an unrestricted research grant from Pfizer (ID 35384245). The institutional review board of the University Medical Center Utrecht (reference number 18-081/c) provided ethical approval and the ethics protocol reference number.

Authors' Contributions

CJJK, CV, and MFP developed the previous e-Exercise programs and used their experience to advise on the development of e-Exercise HA. MAT performed the literature search and the focus group discussions. MAT and MFP created the draft versions of the intervention. MAT made adaptation to the intervention based on the iterative cocreation process. MAT, PdK, and IARK created the draft versions of the information modules (text and videos). All authors contributed to the critical reviewing and editing of the manuscript. All authors approved the final manuscript.

Conflicts of Interest

REGS reports grants from Bayer, NovoNordisk, Sobi, outside the submitted work.

Multimedia Appendix 1
Screenshot of an information module for patients.
[PNG File , 399 KB - jmir_v22i6e16631_app1.png]

References

1. Srivastava A, Brewer AK, Mauser-Bunschoten EP, Key NS, Kitchen S, Llinas A, Treatment Guidelines Working Group on Behalf of The World Federation Of Hemophilia. Guidelines for the management of hemophilia. Haemophilia 2013 Jan;19(1):e1-47. [doi: 10.1111/j.1365-2516.2012.02909.x] [Medline: 22776238]
2. Ahnström J, Berntorp E, Lindvall K, Björkman S. A 6-year follow-up of dosing, coagulation factor levels and bleedings in relation to joint status in the prophylactic treatment of haemophilia. Haemophilia 2004 Nov;10(6):689-697. [doi: 10.1111/j.1365-2516.2004.01036.x] [Medline: 15569163]
3. Stephensen D, Tait RC, Brodie N, Collins P, Cheal R, Keeling D, et al. Changing patterns of bleeding in patients with severe haemophilia A. Haemophilia 2009 Nov;15(6):1210-1214. [doi: 10.1111/j.1365-2516.2008.01876.x] [Medline: 19659938]
4. Roosendaal, van Rinsum, Vianen, van den Berg, Lafeber, Bijlsma. Haemophilic arthropathy resembles degenerative rather than inflammatory joint disease. Histopathology 1999 Feb;34(2):144-153. [doi: 10.1046/j.1365-2559.1999.00608.x] [Medline: 10064394]
5. Timmer MA, Pisters MF, de Kleijn P, de Bie RA, Schutgens REG, Veenhof C. Movement behaviour in adults with haemophilia compared to healthy adults. Haemophilia 2018 May;24(3):445-451. [doi: 10.1111/hae.13441] [Medline: 29493876]
6. Strike K, Mulder K, Michael R. Exercise for haemophilia. Cochrane Database Syst Rev 2016 Dec 19;12:CD011180 [FREE Full text] [doi: 10.1002/14651858.CD011180.pub2] [Medline: 27992070]
7. Lee J, Chang RW, Ehrlich-Jones L, Kwoh CK, Nevitt M, Semanik PA, et al. Sedentary behavior and physical function: objective evidence from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2015 Mar;67(3):366-373 [FREE Full text] [doi: 10.1002/acr.22432] [Medline: 25155652]
8. Qian W, Lam TT, Lam HHW, Li C, Cheung YT. Telehealth Interventions for Improving Self-Management in Patients With Hemophilia: Scoping Review of Clinical Studies. J Med Internet Res 2019 Jul 10;21(7):e12340 [FREE Full text] [doi: 10.2196/12340] [Medline: 31293241]
9. Wagner B, Seuser A, Krüger S, Herzig ML, Hilberg T, Ay C, et al. Establishing an online physical exercise program for people with hemophilia. Wien Klin Wochenschr 2019 Nov;131(21-22):558-566 [FREE Full text] [doi: 10.1007/s00508-019-01548-1] [Medline: 31535221]
10. Bossen D, Kloek C, Snippe HW, Dekker J, de Bakker D, Veenhof C. A Blended Intervention for Patients With Knee and Hip Osteoarthritis in the Physical Therapy Practice: Development and a Pilot Study. JMIR Res Protoc 2016 Feb 24;5(1):e32 [FREE Full text] [doi: 10.2196/resprot.5049] [Medline: 26912378]
11. Kloek CJJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a Blended Physical Therapist Intervention in People With Hip Osteoarthritis, Knee Osteoarthritis, or Both: A Cluster-Randomized Controlled Trial. Phys Ther 2018 Jul 01;98(7):560-570 [FREE Full text] [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
12. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
13. Kloek CJ, Bossen D, de Vries HJ, de Bakker DH, Veenhof C, Dekker J. Physiotherapists' experiences with a blended osteoarthritis intervention: a mixed methods study. Physiother Theory Pract 2020 May;36(5):572-579. [doi: 10.1080/09593985.2018.1489926] [Medline: 29952687]
14. de Vries HJ, Kloek CJ, de Bakker DH, Dekker J, Bossen D, Veenhof C. Determinants of Adherence to the Online Component of a Blended Intervention for Patients with Hip and/or Knee Osteoarthritis: A Mixed Methods Study Embedded in the e-Exercise Trial. Telemed J E Health 2017 Dec;23(12):1002-1010. [doi: 10.1089/tmj.2016.0264] [Medline: 28525310]
15. Sholzberg M, Floros G, Schneiderman JE, Kahr WHA, Rand M, Pluthero F, et al. Effect of moderate intensity exercise on haemostatic capacity in adults with haemophilia A and B: pilot study. Haemophilia 2017 Mar;23(2):e162-e165. [doi: 10.1111/hae.13181] [Medline: 28111839]
16. Krüger S, Weitz C, Runkel B, Hilberg T. Pain sensitivity in patients with haemophilia following moderate aerobic exercise intervention. Haemophilia 2016 Nov;22(6):886-893. [doi: 10.1111/hae.13016] [Medline: 27397488]
17. Runkel B, Von Mackensen S, Hilberg T. RCT - subjective physical performance and quality of life after a 6-month programmed sports therapy (PST) in patients with haemophilia. Haemophilia 2017 Jan;23(1):144-151. [doi: 10.1111/hae.13079] [Medline: 27726259]
18. Cuesta-Barriuso R, Gómez-Conesa A, López-Pina JA. Manual and educational therapy in the treatment of hemophilic arthropathy of the elbow: a randomized pilot study. Orphanet J Rare Dis 2018 Sep 03;13(1):151 [FREE Full text] [doi: 10.1186/s13023-018-0884-5] [Medline: 30176883]

19. Cuesta-Barriuso R, Torres-Ortuño A, Nieto-Munuera J, López-Pina JA. Effectiveness of an Educational Physiotherapy and Therapeutic Exercise Program in Adult Patients With Hemophilia: A Randomized Controlled Trial. Arch Phys Med Rehabil 2017 May;98(5):841-848. [doi: 10.1016/j.apmr.2016.10.014] [Medline: 27865846]
20. Kampshoff C, Peter W, Doormaal M, Knoop J, Meerhoff G. KNGF-richtlijn Artrose heup-knie: verantwoording en toelichting. 2010. Vlieland TPMV URL: https://www.zorginzicht.nl/binaries/content/assets/zorginzicht/kwaliteitsinstrumenten/Verantwoording+KNGF-richtlijn+Artrose+heup-knie.PDF [accessed 2020-05-20]
21. Pulles AE, Mastbergen SC, Schutgens RE, Lafeber FP, van Vulpen LF. Pathophysiology of hemophilic arthropathy and potential targets for therapy. Pharmacol Res 2017 Jan;115:192-199. [doi: 10.1016/j.phrs.2016.11.032] [Medline: 27890816]
22. Timmer MA, Pisters MF, de Kleijn P, Veenhof C, Laros-van Gorkom BAP, Kruip MJHA, et al. How do patients and professionals differentiate between intra-articular joint bleeds and acute flare-ups of arthropathy in patients with haemophilia? Haemophilia 2016 May;22(3):368-373. [doi: 10.1111/hae.12858] [Medline: 26634961]
23. Timmer MA, Pisters MF, de Kleijn P, de Bie RA, Fischer K, Schutgens RE. Differentiating between signs of intra-articular joint bleeding and chronic arthropathy in haemophilia: a narrative review of the literature. Haemophilia 2015 May;21(3):289-296. [doi: 10.1111/hae.12667] [Medline: 25854291]
24. Hildebrandt V, Bernaards C, Hofstetter H. TNO. 2015. Report Movement and Health. In Dutch: Trendrapport bewegen en gezondheid 2000/2004 URL: https://www.kenniscentrumsportenbewegen.nl/kennisbank/publicaties/?kb_id=16038 [accessed 2020-05-20]
25. Pisters MF, Veenhof C, de Bakker DH, Schellevis FG, Dekker J. Behavioural graded activity results in better exercise adherence and more physical activity than usual care in people with osteoarthritis: a cluster-randomised trial. Journal of Physiotherapy 2010;56(1):41-47. [doi: 10.1016/s1836-9553(10)70053-9]
26. Kulkarni R. Use of telehealth in the delivery of comprehensive care for patients with haemophilia and other inherited bleeding disorders. Haemophilia 2018 Jan;24(1):33-42. [doi: 10.1111/hae.13364] [Medline: 29205699]
27. Pietrzak E, Cotea C, Pullman S, Nasveld P. Self-management and rehabilitation in osteoarthritis: is there a place for internet-based interventions? Telemed J E Health 2013 Oct;19(10):800-805. [doi: 10.1089/tmj.2012.0288] [Medline: 24003979]
28. Kloek C, van Tilburg M, Staal J, Veenhof C, Bossen D. Development and proof of concept of a blended physiotherapeutic intervention for patients with non-specific low back pain. Physiotherapy 2019 Dec;105(4):483-491 [FREE Full text] [doi: 10.1016/j.physio.2018.12.006] [Medline: 31031023]
29. May CR, Finch TL, Cornford J, Exley C, Gately C, Kirk S, et al. Integrating telecare for chronic disease management in the community: what needs to be done? BMC Health Serv Res 2011 May 27;11:131 [FREE Full text] [doi: 10.1186/1472-6963-11-131] [Medline: 21619596]

Abbreviations

CeHRes: Center for eHealth Research
eHealth: electronic health
HA: hemophilic arthropathy
OA: osteoarthritis

Edited by G Eysenbach; submitted 09.10.19; peer-reviewed by D Bossen, A Wijnen, H de Vries; comments to author 12.11.19; revised version received 16.12.19; accepted 26.01.20; published 19.06.20.

Please cite as:
Timmer MA, Kloek CJJ, de Kleijn P, Kuijlaars IAR, Schutgens REG, Veenhof C, Pisters MF
A Blended Physiotherapy Intervention for Persons With Hemophilic Arthropathy: Development Study
J Med Internet Res 2020;22(6):e16631
URL: http://www.jmir.org/2020/6/e16631/
doi:10.2196/16631
PMID:32558654

Original Paper

Young People's Experience of a Long-Term Social Media–Based Intervention for First-Episode Psychosis: Qualitative Analysis

Lee Valentine[1,2], BA, MSocWk; Carla McEnery[1,2,3], BPsychSc (Hons), PhD; Shaunagh O'Sullivan[1,2], BA (Hons), MPsychSc; John Gleeson[4], BA (Hons), MPsych (Clin), PhD; Sarah Bendall[1,2*], BA, PGDipClinPsych, MA, PhD; Mario Alvarez-Jimenez[1,2*], BSc (Hons), DClinPsy, MAResearchMeth, PhD

[1]Orygen, Parkville, Australia
[2]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia
[3]Centre for Mental Health, Swinburne University of Technology, Hawthorn, Australia
[4]School of Behavioural and Health Sciences, Australian Catholic University, Melbourne, Australia
[*]these authors contributed equally

Corresponding Author:
Lee Valentine, BA, MSocWk
Orygen
35 Poplar Rd
Parkville, 3052
Australia
Phone: 61 0417398175
Email: lee.valentine@orygen.org.au

Abstract

Background: Digital mental health interventions present a unique opportunity to address the lack of social connection and loneliness experienced by young people with first-episode psychosis (FEP). The first generation of digital interventions, however, is associated with high attrition rates. Social media presents an opportunity to target this issue. A new generation of digital intervention has harnessed the popularity of social media to both promote engagement and foster social connectedness in youth mental health interventions. Despite their potential, little is known about how young people engage with, and experience, social media–based interventions as well as the optimal design, implementation, and management needed to ensure young people with psychosis receive benefit.

Objective: This study aimed to explore how young people engage with, and experience, a long-term social media–based mental health intervention designed to address social functioning in individuals with FEP.

Methods: This qualitative study was based on 12 interviews with young people who used Horyzons, a long-term social media–based mental health intervention, as part of a previous randomized controlled trial. A semistructured phenomenological interview guide with open-ended questions was used to explore young people's subjective experience of the intervention. All interviews were recorded and transcribed verbatim. Data were analyzed using interpretative phenomenological analysis.

Results: A total of 4 superordinate themes emerged during the analysis including (1) shared experience as the catalyst for a cocreated social space, (2) the power of peer support, (3) an upbeat environment, and (4) experiences that interrupt being in Horyzons.

Conclusions: We found that Horyzon's therapeutic social network fostered a connection and an understanding among young people. It also aided in the creation of an embodied experience that afforded young people with FEP a sense of self-recognition and belonging over the long term. However, although we found that most young people had strong positive experiences of a social connection on Horyzons, we also found that they experienced significant barriers that could substantively interrupt their ability to use the platform. We found that social anxiety, paranoia, internalized stigma, lack of autonomy, and social protocol confusion interfered with young people's usage of the platform. From a design perspective, digital interventions are flexible and thus equipped to begin addressing these implications by providing customizable and personalized treatment options that account for varying levels of social connection and psychological need that could otherwise interrupt young people's usage of social media–based interventions.

(J Med Internet Res 2020;22(6):e17570) doi:10.2196/17570

KEYWORDS

social media; social networking; youth; young adult; psychotic disorders; mHealth; qualitative research

Introduction

Background

The current treatment of psychosis focuses primarily on symptom reduction [1]. Psychological and pharmacological interventions are effective in treating symptoms and preventing relapse to a degree; however, many young people with first-episode psychosis (FEP) experience significant difficulties with regard to social functioning and report high levels of social isolation [2-4]. Despite these findings, very few studies have assessed interventions targeting social functioning as a primary outcome [2,3].

Even in the earliest stages of the disorder, young people with FEP experience reduced social opportunities, have limited social networks, report difficulty developing and maintaining social relationships, and experience high levels of loneliness [1,3,5]. These findings have led to an increased focus on effective interventions that target loneliness and social connection [6]. However, despite the recognized importance of including social connection within a broader recovery framework, there are limited evidence-based solutions available to address this need [3,7,8].

Digital mental health interventions present a unique opportunity to address a lack of social connection and loneliness in FEP [3,8-14]. However, the first generation of digital interventions is associated with high attrition rates [15,16]. Social media presents an opportunity to target this issue. A new generation of digital interventions has harnessed the popularity of social media to both promote engagement and foster social connectedness [8,9,12,13].

Objective

Despite their potential, little is known about how young people engage with, and experience, social media–based interventions as well as the optimal design, implementation, and management needed to ensure young people with psychosis receive benefit. In addition, little is known about how these interventions are experienced over the long term. This study will address this gap by exploring young people's experiences of a long-term social media–based intervention to inform the emerging generation of digital mental health interventions. Therefore, this study aimed to explore how young people engage with, and experience, a long-term social media–based intervention designed to address recovery in FEP.

Methods

Setting and Design

This phenomenological qualitative study was based on interviews conducted between October 2018 and March 2019. Young people interviewed in this study had previously participated in the Horyzons randomized controlled trial (RCT; ACTRN12614000009617). The Horyzons RCT evaluated whether a long-term 18-month moderated online social therapy (MOST) platform, known as Horyzons, was superior to 18 months of regular care, following discharge from 2 years of treatment at an FEP early intervention service [8]. The MOST platform was co-designed and developed by a multidisciplinary team of peer workers, clinical psychologists, computing and information systems researchers, software developers, illustrators, and writers [17]. Horyzons was underpinned by a strengths-based and positive psychology framework designed to generate long-term functional recovery and increase well-being in FEP. It was a closed digital platform with options to engage in web-based therapy pathways (Figure 1), speak with peer and clinical moderators (both privately via a web-based chat and publicly on the social network), and connect with other young people via the therapeutic social network, which was referred to as the *newsfeed* (Figure 2) [8]. The digital functions of the platform, such as the web-based therapy pathways, were reinforced by human support provided by the clinical moderators through a supportive accountability model [18]. This contact was primarily provided through web-based chats and phone calls. A key clinical moderator was also allocated to each participant and was responsible for overseeing their engagement and progress over the span of the intervention [18]. Young people could connect on the social network by posting and commenting, sharing experiences, and giving and receiving support. Conversation could occur between participants in a freeform way on the newsfeed and they could also take part in focused group discussions through a feature referred to as *Talk It Outs* [8]. These discussions were overseen by peer moderators. As described in the Horyzons protocol, the functions of the platform were "designed to reinforce each other, creating a flow for the young person between the social and therapy elements" [8]. For the Horyzons RCT, a total of 170 young people approaching or recently discharged from the Early Psychosis Prevention and Intervention Centre in Melbourne, Australia, were recruited. Of the 170 participants, 85 were randomly allocated to the specialist treatment group. This gave them access to the Horyzons platform for an intervention period of 18 months. The remaining 85 participants continued with treatment as usual.

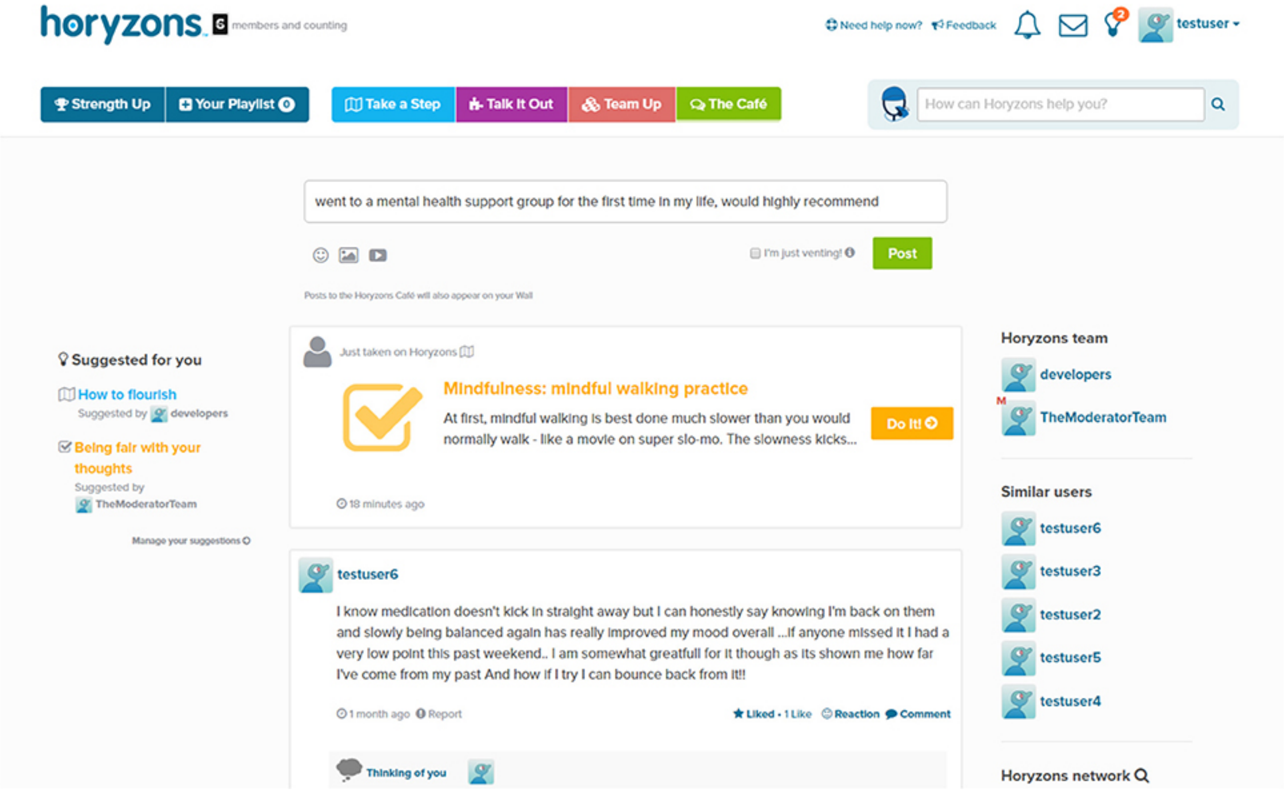
Figure 1. Everyday mindfulness therapy step.

Everyday mindfulness

In this Step

- Mindfulness is a kind of meditation that you don't need to be in one place or time for. It's great for everyday moments and can be done in a few seconds or a few minutes.
- Once you start practicing tuning in and anchoring yourself with your body, breath and senses, you can use those skills in everyday moments to help stay present, aware and alive to the world.

Figure 2. The Horyzons news feed.



went to a mental health support group for the first time in my life, would highly recommend

Mindfulness: mindful walking practice

At first, mindful walking is best done much slower than you would normally walk - like a movie on super slo-mo. The slowness kicks...

testuser6

I know medication doesn't kick in straight away but I can honestly say knowing I'm back on them and slowly being balanced again has really improved my mood overall ...if anyone missed it I had a very low point this past weekend.. I am somewhat greatfull for it though as its shown me how far I've come from my past And how if I try I can bounce back from it!!

Participants

The study inclusion criteria were as follows: (1) participation in the Horyzons RCT, (2) randomization to the treatment group of the Horyzons RCT, and (3) completing the Horyzons RCT. With the aim of sampling a range of young people with different levels of platform usage [19], a quantitative categorization of Horyzons participants was conducted. This categorization was based on each participant's overall number of log-ins to the Horyzons platform. Thus, all 85 participants randomized to the treatment group of the Horyzons RCT were divided into 4 quartiles based on the number of overall log-ins to the platform over the 18-month intervention period. Overall, log-ins ranged from 1 to 1568. The number of log-ins associated with each usage group is detailed in Table 1. The quartiles represented 4 usage groups: very low, low, moderate, and high usage. Young people from each usage group were then randomly contacted via phone call or text and invited to participate in a qualitative interview.

Table 1. Log-ins, mean, and standard deviation associated with each usage group.

Usage group	Total (N=84), n (%)	Log-ins	Mean (SD)
Very low	21 (25)	1-8	5 (2.2)
Low	21 (25)	9-23	15 (4.9)
Moderate	22 (26)	24-80	31 (19.3)
High	20 (24)	88-1569	366 (405.7)

Data Collection

All interviews were conducted using a phenomenological approach [20]. This interview method is concerned with the lived experience of a phenomenon, as experienced subjectively by the individual [21]. The interview schedule comprised 2 extremely broad, open-ended questions: (1) What was Horyzons? and (2) What was Horyzons like? These questions were designed to elicit young people's experiences of the Horyzons platform without leading them in a particular direction. Subsequent prompts were used if necessary; however, the method calls for the interviewer to follow the leads presented by the participant. Participants were given the option of completing the interview at the Orygen clinic, home, or in a public place (eg, library or café). Subsequently, interviews were conducted in a range of locations. Before beginning each interview, participants were required to complete a plain language consent form and given the opportunity to ask questions. Author LV conducted all interviews and had established relationships with all participants due to a previous position as a research assistant on the Horyzons RCT. LV is an early career researcher with extensive quantitative interviewing experience and is currently undertaking a research project using phenomenological methodological interviewing. Field notes were taken during the interviews, and a reflective log was maintained over the interview period. In accordance with the interpretative phenomenological analysis (IPA) approach, data collection, transcription, and preliminary analysis were conducted in parallel [22]. All interviews were audio-recorded and transcribed verbatim. All participants were reimbursed Aus $20 (US $12.63) for their participation in the study. All young people approached to participate in the study were interviewed with the exception of one young person who became distressed at the beginning of the interview due to acute psychotic symptoms, and the interview was quickly terminated. The recruitment of participants continued until a saturation of themes was reached at 11 participants [20]. The twelfth participant was then sampled to ensure that no additional themes emerged [20].

Data Analysis

Interviews were analyzed using IPA. In accordance with IPA, analysis for the study commenced once the first qualitative interview had been completed and transcribed [21]. Interviews were analyzed individually and in relation to all other interviews completed over the course of the study in a multilayered and overlapping IPA framework. Themes were derived directly from the data collected. Given that interpretation can be based on the context in which the phenomenon is viewed, member checking was not offered to participants because of the interpretative nature of the analysis used [22]. To begin, multiple close readings of each transcription were conducted. Initial responses to the data were noted and recorded. Codes or brief summaries describing the data were recorded. These codes formed the bases of emerging themes and patterns across the overall data set. The identification of reoccurring themes within and across interviews emerged and was noted. Some themes were distinct from one another, whereas others merged as superordinate themes. Strategies to maintain rigor, reliability, and validity throughout the analysis process included prolonged engagement with the data and thick and rich descriptions of results [23]. Author LV conducted the analysis and engaged in regular reflexive discussion regarding the codes and subsequent themes with the senior author (SB) [23]. Microsoft Excel was used in the data analysis process.

Results

A total of 12 participants aged 19 to 28 years (mean 23 years) were included in the study; 7 of the participants were female and 5 were male. All of the female participants were cisgender. Of the male participants, 4 were cisgender and 1 was transgender. Multiple participants from each of the 4 usage groups were interviewed. Interview times ranged from 30 to 86 min. The time period in which participants were interviewed for this study following the last date they were able to access the Horyzons platform ranged from 1 to 5 months. The mean follow-up time was 3 months. All participant names were replaced with an alias to ensure privacy. Table 2 gives a description of participant characteristics.

In total, 4 superordinate themes emerged during the analysis, with each theme divided into subthemes to comprehensively capture the phenomenon explored. The 4 superordinate themes were (1) shared experience as the catalyst for a cocreated social space, (2) the power of peer support, (3) an *upbeat* environment, and (4) interruptions to *being in* the Horyzons space.

Table 2. Descriptive summary of participant usage groups, alias, gender, age, follow-up time, length of interview, and overall log-ins to the Horyzons platform.

Usage group and alias	Gender	Age (years)	Time since last possible log-in (months)	Interview length (mins)	Overall log-ins, n
High					
Sam	M[a]	28	2	49	1568
Babak	M	23	3	25	109
Audrey	F[b]	24	5	20	1230
Moderate					
Emily	F	26	4	49	66
Jacob	M	21	5	50	65
Low					
Tom	M	22	1	86	18
Celeste	F	23	3	58	12
Olivia	F	22	4	79	13
Abigail	F	19	5	32	23
Very low					
Linh	F	25	2	56	6
Eugene	M	25	2	42	8
Ruby	F	20	3	45	7

[a] M: male

[b] F: female.

Shared Experiences as Catalysts for a Cocreated Social Space

Horyzons Was a Place and "I Felt at Home"

Young people used language commonly associated with the description of physical places when describing experiences of the Horyzons platform. That is to say that they spoke of Horyzons as a tangible or an embodied space that they could *arrive at* and *be in*. For instance, Audrey's description of sharing accomplishments with her peers on the therapeutic social network evokes imagery of *traveling to* and *being at* a physical destination:

> *So, I went there, and I would talk, talk about my achievements. I felt proud of myself for doing something. People would encourage me or commend me, yeah. So, it was a nice place to be.* [Audrey]

Audrey uses the spatial and social descriptors "went there," "talk about," "people would encourage me," and "a nice place to be" to describe moments that are technically confined to a computer, tablet, or phone. However, her phrasing of these experiences points toward a perception of Horyzons as a physical experience. Instead of detailing how she *uses* the platform, Audrey describes an experience of actually *going there* and *being in* the Horyzons social space.

"Don't Necessarily Know Each Other Personally" But a Connection Through Shared Experience

A number of participants described an experience of *knowing* other members of the platform. Interestingly, however, the *knowing* in most cases did not equate to a direct social relationship or connection, either physically or digitally, but was instead based on the participants' *common* history of mental health experiences:

> *It's like… war veterans… where you have all the gatherings of people who have been through wars and they don't necessarily know each other personally, but they've all got that common war experience. And they can relate to each other and they know what they've been through.* [Sam]

Sam's description builds upon commonplace social definitions of *knowing* and *connection.* It offers an alternative construct of what it might be to *know* other members in the particular place that is the *Horyzons space*. In this space, *knowing* each other may not be contingent on direct social contact but instead on the concept of *understanding* and *relating* to the *common* experience of being a young person with lived mental health experience.

"I Just Felt Like Relieved That I Wasn't The Only One"

Young people detailed a range of positive emotional responses at not being *the only one* with an experience of psychosis and other mental health experiences on the Horyzons platform.

Being in this space with others who shared this *common* history had a strong impact on participants. Many expressed that they could *relate* to others and both *understand* and feel *understood* in this environment. The sharing of experiences on the Horyzons platform by group members had a normalizing impact for some participants. For instance, Sam expressed:

> *It helped sort of reinforce that some of my experiences and feelings are normal… I'm not a weirdo.* [Sam]

Some members found it easier to share their own experiences with the group given they were in a shared social space with others who also had a lived experience of mental ill health:

> *My story is not as twisted as it is common, shit that happens. I could feel more comfortable sharing knowing that it's common...* [Jacob]

For some, sharing their personal stories with the group was a therapeutic activity in itself. The act of sharing supported the young person to unpack and reflect on their mental health experiences:

> *I wanted to support them, so I think I reflected on what I have been through…Actually gave them my experience, how I fought through it.* [Emily]

For Olivia, the shared experience of mental ill health sparked a sense of pride, pride that the group existed despite the adversity that its members face:

> *For me, it was just really important to see that there were other people that were kind of just saying like,' yeah, I'm proud.'… I think it's really important to be proud even though you've got fucking mental problems.* [Olivia]

Some participants identified the benefit of learning about other people's experiences, even if their stories were not directly relevant to the young person's own life. This information was a way to feel connected to others, receive encouragement, and access a diverse range of coping methods:

> *We didn't always have the same illness or anything. It was nice to hear people, people's experiences. It was encouraging to hear, yeah...Um, just how they dealt with the problem...And everyone approached them differently.* [Abigail]

One particular participant identified other members on the Horyzons platform that he knew from offline spaces. For him, this social overlap helped to reinforce that "mental health is a common thing."

For some participants, the opportunity to share a space with peers that they were able to relate to, and understand their experiences, reminded them of painful memories when their mental health struggles had been minimized, misunderstood, or mocked:

> *When I was with Orygen I told them the first time what had happened [family], and they're like "nah it's not true, it's just all in your head, it's all in your mind", and they didn't believe me, and it kinda just felt like I couldn't trust them.* [Emily]

"A Barrier That's Automatically Broken Down"

As described in the themes above, the cocreation of space and the concept of a *common* mental health history came together to create a social environment on Horyzons that participants described as *more intimate* than other spaces. Many participants described a greater sense of ease on the platform between peers than in other social spaces, both on the web and offline. This ease between young people acted to break down conventional social barriers and accelerated bonding:

> *Whereas I guess with traditional social media you don't just go up to random people and say "hey I've got this thing, do you want to talk about it?" whereas on Horyzons it was sort of specially crafted for that sort of thing, so it was a lot easier to approach people and sort of form bonds and stuff.* [Sam]

For participants, the awareness that they were among others with a shared experience meant they felt less judged and as a consequence of this, more relaxed to share and connect:

> *I connected to them a lot more easier than I would with others… I felt like I was in a more calm relaxed space connecting with them than other people that I meet at Uni… Like they were more open, non-judgemental... I felt relaxed…Yeah I felt really relaxed.* [Emily]

This dynamic appears to be somewhat cyclical in nature for participants. The recognition that their peers have a shared experience and a *more open* and *nonjudgmental* approach results in a more *relaxed* environment conducive to sharing. Once this process has occurred, several young people described an experience whereby they saw others sharing and this created a sense of positive responsibility to their peers in which they described wanting to "own up a bit more" to the group:

> *They were pretty much open with what they were going through…and I thought to myself 'I should open up at least and see where I would go'…Just like take a shot at it.* [Emily]

The Power of Peer Support

"Don't Really Know Unless You've Experienced It"

In the Horyzons context, the function of offering and receiving support was generally operationalized through public posts and replies on the social network. Participants expressed that offering and receiving advice and support from others with lived mental health experience, otherwise known as peer support, was a valuable practice that is unique to the peer-to-peer relationship. This dynamic is situated within the context of shared mental health history, a history that a clinician may not always have direct experience with. In the words of Abigail:

> *They don't really know unless [they've] experienced it.* [Abigail]
>
> *Like peer support really is helpful for me in particular, just cos like, I kinda find it more difficult to open up to someone that doesn't really fully get the picture, even if they're like a completely trained professional.* [Ruby]

Participants communicated that their life experiences should position them at the center of the Horyzons platform. They identified that their mental health journey can help to guide current day problem solving for themselves and others:

> *Apart from starting it and summarizing it at the end [group discussions], I feel like it was mostly handed over to the users, which is probably a good thing*

> *because it allows the people with the experiences to sort of come up with their own ideas and thoughts.* [Sam]

For some young people, the advice of a peer is elevated above that of a "completely trained professional" or "someone giving a TED Talk." Some young people expressed that support from a person with lived experience would be easier to implement in their own life because that individual would "fully get the picture" and be more likely to offer "practical" and "realistic" advice:

> *I want to know what real everyday people do. Not a social media presence person, just the person on the other side of the road. What do they do? ...It might not work for me…but it's probably most interesting because it's probably most um, uh, doable?* [Celeste]

"I Love Being a Massive Resource for That Sort of Thing"

Many participants described offering support and advice to others based on knowledge gained via their own mental health experiences as an empowering, validating, and, at times, joyful function of the Horyzons platform. Emily expressed, "I felt really calm, relaxed and over the moon because I was able to help someone."

"Responsibility to… Basically Just be Their Secondary Psychologist"

In total, 2 participants expressed strong reservations regarding what they experienced as an unwanted obligation to provide peer support to other members on the platform. Ruby stated, "just that fear if someone messages me, I have a responsibility to talk to them, to support them, to basically just be their secondary psychologist."

Ruby describes experiencing a lack of control over the peer support function of the platform, "I don't always like it when people open up to me when I don't ask for it." As well as anticipatory fear that denying a request for peer support from another member would result in a negative interaction, "I was scared that they'd like start cyber-stalking me or something."

An "Upbeat" Environment

As mentioned earlier, a positive psychology framework underpinned the Horyzons RCT. This framework informed how clinicians moderated the platform. For example, using a strengths-based reframing approach to young people's posts. Participants reported a spectrum of emotional responses to the environment that this clinical approach created on the Horyzons social network.

On one end of the spectrum, Celeste described feeling frustrated and patronized by this approach and described the social network as "mind-numbingly positive." She argued that this type of environment impedes the process of sharing the "bad stuff" that young people living with mental health problems can experience. To discuss difficult topics, which Celeste identified as an important process of mental health recovery, members would need to go against the grain of the established "upbeat" environment.

Celeste described feeling alienated by this approach and went so far as to describe a phenomenon akin to an uncanny valley experience. That is to say, the platform looked similar to other social networks, but that the environment did not feel like any other social spaces she was accustomed to. This irregularity led to a feeling of unease:

> *It was just odd, and it didn't feel like it could be genuine because it's not my experience of how people interact with each other in the everyday world, you know? [Celeste]*

In contrast to Celeste's experience, Audrey identified the "positive" environment on Horyzons as an important element of her high usage level. She did not want to spend time at a place where "the people are all negative."

Olivia's experience was similar to Celeste's in that she experienced the "upbeat" nature of the platform as different from that of commercial social network sites (SNS). However, she also identified that although it was positive for people with mental health difficulties to connect with others in certain mental health pockets of social media, such as "Instagram," certain relationships and dynamics that are borne from these unregulated digital spaces can have negative impacts on the young people who frequent them. For example:

> *You know, go on Instagram-... type in eating disorder, you got a whole world there. You know? You can type in depression; you've got a whole world of people struggling. And it's a good and a bad thing. So, it's good as in they can connect with other people. But, it's bad as in you get triggered constantly. Rather than that, Horyzons was, "Yeah, we're all struggling... but this is a place for improvement."* [Olivia]

For Eugene, the experience of being in an environment with a psychological lens tarnished his experience of the social network. He said:

> *…that's what my brain tells me, it's just like, super wanky. Like, I think psychology and… that sort of stuff is really wanky.* [Eugene]

He did, however, describe experiencing something of a psychological shift when the platform was able to support him during a time when he was experiencing "major issues." As a result, he identified that Horyzons "may have a point" and may not be "so wanky." A shift, such as a change in mind or need, was a phenomenon reported by other participants in reference to a range of issues on the platform.

Experiences That Interrupt Being in the Horyzons Space

Participants identified a lack of motivation to participate in the social network as well as a range of experiences that interrupted their ability to *be in* the Horyzons social space in an immersed or *relaxed* way. These interruptions gave way to a self-conscious or difficult moment, or a series of moments, which created tension between the young person and their ability to use the platform. Participants identified that these incidents could occur before, during, or after single sessions and throughout the overall intervention experience itself. Some participants identified multiple interruptions, or ruptures, throughout their Horyzons experience.

Absence of Motivation

A number of participants described an absence of motivation to participate in the social network, either before any initial engagement with the platform or at some point during the intervention period. This was due to the personal assessment that the social network was not relevant to them and their needs.

For instance, Tom identified that from the beginning of his intervention period, he intended to engage solely with the psychosocial content of the platform and expressed an active disinterest in the social network. He described experiencing ample levels of social media in his daily life and was not interested in getting "bogged down" in that aspect of the platform.

Jacob identified that although Horyzons was valuable during the first half of his intervention period, he went on to experience a "tipping point" in which he felt he had moved beyond what the platform was able to offer him. He explained, "that's not the stage of life that I'm at anymore."

Experiences That Interrupted Being in the Horyzons Space

A number of young people identified paranoia, social anxiety, and internalized stigma as experiences that interrupted them from *being in* the Horyzons space. These types of interruptions generally occurred *during* the young person's engagement with the platform. However, while they were attempting to engage, an emotional or psychological barrier arose that impeded that engagement. This ultimately took them out of the immersed or *relaxed* experience of *being in* Horyzons.

Paranoia in relation to Horyzons itself and paranoia in relation to social media use more broadly was a phenomenon described by 2 participants. Emily described experiencing paranoia in relation to social media use more generally, whereas Olivia expressed experiencing disruptive paranoia in regard to the Horyzons platform directly. For instance, she said, "I just get really paranoid about like who's behind this and stuff like that…like what are the motives behind this, and stuff like that." This experience of paranoia interrupted Olivia's capacity to share candid information and engage directly with other members on the platform. The act of even logging in to the site was difficult, as she explained, "They're watching me too closely. They know too much."

Participants also described experiencing social anxiety related to the complex psychological process of posting in the social network. This anxiety is generally related to how others might perceive them and prevented many young people from posting on the platform. For example, Ruby expressed fear of overwhelming others with the content of her posts and wrote and deleted "5 to 10" messages that she ultimately never shared with the group. Babak felt that his thoughts may not be important enough to share. A number of participants expressed uncertainty about Horyzon's general social protocol and speculated that they were concerned about breaking "weird, unwritten rules" that they were not aware of. Multiple participants described feeling nervous that if they were to post, then other members may not respond, or may not respond in a desired way. For instance:

> *It would just sit there, and no one would respond to it...and then I'd be like "ah ok" I'm just kind of in an echo chamber [laughs]...so embarrassing … [interviewer: how come?] cos like no one would really care. It would just kind of sit there and look weird. It's like having a spotlight on yourself and everyone just staring at you and it's completely silent.* [Ruby]

This idea of *sharing anxiety*—that is, the phenomenon of experiencing anxiety at the prospect of posting a message to the social network—interrupted young people's ability to act spontaneously in the social space. They reported feeling frozen in moments of rumination about possible negative consequences of posting that could go on for a short time, such as minutes, or could impede them from posting on the platform throughout their whole Horyzons experience, despite identifying that there was content they would like to share with other members. Interestingly, however, all but one participant who reported experiencing sharing anxiety about posting but went on to post did not go on to report any negative consequences of posting. On the contrary, Audrey reported that although replies to her posts in the Horyzons space were slower than she was accustom to on other social networks, she described feeling "patient" while waiting a "day or two" for a reply and felt "excited" when she received a response. Furthermore, she felt "happy" to receive fewer comments on Horyzons than compared with other social media because she considered the replies received on Horyzons to be genuine and "not just for attention."

However, one particular participant expressed that he continued to think about a post he shared on the network many months after he originally posted it. After initially sharing the post, Jacob became concerned that a moderator had potentially "hidden" it from the public feed because of the possibility that it was inappropriate:

> *I don't know if this is true or not, but I don't know if a moderator actually hid the post for a bit… I'm not sure… I didn't write again, because maybe then I started second thinking the post, maybe it wasn't appropriate.* [Jacob]

This confusion transformed into anxious rumination that interrupted Jacob's ability to post again during his intervention period. This may also be related to his decision that he had reached a "tipping point" and no longer needed the site.

Eugene experienced engaging with the Horyzons platform as incongruous with his sense of self. He considered himself older, more independent, and more advanced in his career than the other members of the platform and as a result, he felt an internalized pressure that he "shouldn't need this." Eugene's social struggle to engage with the platform was also closely tied to the judgments he was concerned his colleagues would make of him. He was concerned that they would not "trust" him professionally if they were aware of his mental health history. Eugene appeared to internalize this stigma and his ability to access support, even when identified as needed, was interrupted.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to qualitatively explore young people's experiences of using a long-term social media–based intervention in FEP recovery. In relation to social connection, we found that experiences could be grouped into 4 superordinate themes, as detailed above. These themes have the potential to inform the development and implementation of social media–based interventions. We found that (1) peer support on a therapeutic social network could be beneficial and engaging for some, and burdensome for others; (2) the number of log-ins did not neatly reflect a young person's experience of the intervention, (3) unclear social protocol created an uncomfortable social environment in the digital space, and (4) social anxiety, paranoia, internalized stigma, and the perception of limited autonomy could interrupt a young person's ability to engage with the platform.

In line with previous findings, results from this study indicated that many participants lacked a confidant in their everyday life, a friend with whom they could confide in regarding their mental health experiences [1]. We found that the therapeutic social network offered participants a much-needed opportunity to feel heard, understood, and supported by their peers. For some, this was the first time they had felt connected to others in this way. These findings support previous research that has identified the value of peer support practice in mental health [24,25]. The supportive dynamic between peers on Horyzons aided in the creation of an embodied space in which young people experienced an ongoing sense of belonging. This study found that young people place great value on the advice and support offered by their peers, at times above that of a trained mental health professional. Some advised that they were more likely to activate behaviors suggested by their peers than those of a clinician. This example emphasizes the powerful role that peer support can, and does, play in shaping young people's mental health recovery. It also highlights the instrumental way peer support can create a digital environment that is (1) an engaging pathway to mental health treatment and to the web-based intervention itself and (2) a potentially therapeutic dynamic in itself.

In addition, this research also found that it was possible for young people to experience anxiety, to experience burden, and to perform avoidant behaviors at the perceived expectation to provide peer support to others on the web. This example illustrates the varied needs among young people (ie, what some find helpful, others may find burdensome), as well as the need to consider the autonomy of the user when designing and implementing interventions. This finding highlights the significance of recognizing young people as individuals with individual needs. Fortunately, the mode of digital intervention is well equipped to address this implication by providing customizable treatment options that are applicable to very large groups of diverse individuals.

We found that young people have varying levels of interest with regard to social connection on the web. Therefore, designing customizable interventions that allow young people to identify the importance of this function in regard to their own mental health treatment journey is recommended. For example, personalized onboarding options empower young people to determine the relevance of digital intervention functions, such as the level of peer support they wish to engage with. This places more power in the hands of the user and, in accordance with self-determination theory [26], has the potential to strengthen their engagement with the intervention overall. Furthermore, personalization should be temporally flexible, as this study also found that young people's mental health needs were dynamic in nature and could change over the course of a long-term intervention.

The knowledge that young people had internal experiences that interrupted their ability to be *in* the Horyzons space was a novel finding that has not been previously explored. By asking young people about their experiences of using the platform, we were able to elicit internal barriers to use. Social anxiety, level of paranoia, internalized stigma, and lack of perceived autonomy in providing peer support actively interrupted young people's ability to remain in the Horyzons social space. In line with previous findings that social anxiety is a significant health issue for young people experiencing FEP, this study found that many participants described experiencing social anxiety when using the Horyzons platform [11]. We suggest that anticipating and accounting for these experiences in the design and implementation phase of a digital platform could potentially enhance a young person's ability to use social media–based interventions. This study found that experiences of social anxiety in the digital therapeutic space may differ from that of offline social anxiety, or even anxiety in commercial social media spaces, such as Instagram and Twitter. For example, fear of overwhelming peers with the content of a mental health post may be unique to the therapeutic social network space. As such, young people may experience a new phenomenon of *therapeutic social media anxiety*. As such, anticipating and attempting to circumvent these experiences may increase young people's level of engagement and/or benefit from social media–based interventions. On the basis of these findings, we suggest 3 tangible approaches to addressing digital social anxiety in the therapeutic social network space: (1) As discussed above regarding the personalization of web-based peer support, offer young people the power to customize their level of engagement with others. (2) Pre-empt and creatively account for prevalent digital social anxieties in the design phase of an intervention. For instance, work to normalize the experience of posting anxiety by signposting this as a common fear through comics and quotes from previous users, this may work to normalize and reduce hesitations experienced by the user. (3) Future research should focus on the design of a tool that measures these types of engagement interruptions.

In line with previous findings [27], the results from this study indicate that the level of usage, in this case, the number of log-ins, does not neatly align with a young person's level of expressed connection with the platform. For instance, young people categorized as low or very low users still contributed to themes that characterized Horyzons as a supportive, connected, and embodied experience. The more limited use of the platform appeared to be driven less by indifference or dislike for the intervention itself and more by young people's complex and dynamic social and psychological experiences that interrupted their use of the platform. For example, social anxiety, internalized stigma, level of paranoia, and previous experiences providing peer support were some of the overarching themes that interrupted usage and were related to lower log-ins on the platform. Most young people had strong positive experiences of social connection on Horyzons. Importantly, however, the majority also experienced barriers to engaging with the platform. That is to say, the helpfulness of Horyzons was considered high, but so were the barriers.

This study found that designing a therapeutic social media–based intervention, such as Horyzons, in a similar way to existing commercial SNS archetypes, such as Facebook or Instagram, could create a sense of unease or discomfort for participants due to conflicting social protocols. For example, in many ways, Horyzons felt similar to commercial social media, but there were also distinct differences felt by participants with regard to the different social norms of the platform. For instance, Horyzons encouraged supportive engagement between members. By comparison, a similar act of responding to a personal post on a stranger's page on Facebook could be considered a social transgression. This highlights a nuanced misalignment between therapeutic and commercial social media protocols. In other words, Horyzons looked and functioned similarly to Facebook, which for participants evoked a Facebook-like behavioral framework. However, the social protocol encouraged by Horyzons did not exactly reflect Facebook's social protocol. As a result, young people reported feeling concerned or anxious about performing suggested tasks on Horyzons for fear of behaving inappropriately socially. This finding is in line with previously identified concerns [28] that digital mental health services are at risk of engaging in *psychological skeuomorphs* or the danger of designing derivative models of care and in doing so miss the opportunities that come with reimagining virtual models of care. In some ways, that Horyzons felt similar to commercial SNS may have led to higher user engagement because it is popular and compelling to young people. However, it may have also led to young people experiencing social anxiety or feeling socially exposed due to the nature of the content (ie, sharing mental health experiences in an arena that is not typically experienced in this way). Reimagining how social media–based interventions could still be compelling for young people due to their popularity but feel safer or more socially comfortable to participate in these spaces could be a possible way to reduce internal barriers to engagement. Thus, further research on both features and measures that can enhance engagement or reduce interruptions to a young person's experience of social media–based interventions over the long term would be of interest.

Strengths and Limitations

Recruiting a large sample for the original Horyzons RCT contributed to wide variation in platform usage. Thus, for this study, we were able to purposefully sample a diverse representation of young people from a variety of usage groups with the aim of exploring different user experiences that may have been associated with the level of usage. Although we found that low usage did not result from a blanket disinterest in the platform, this variety in sampling allowed for a better insight into the internal experiences that both drove and hindered usage on the platform from a variety of users. Additionally, this is the first study to explore an 18-month intervention period and as such afforded us the opportunity to explore adherence, experiences, and internal barriers to use over time, which had not yet been previously explored for this length of duration. With regard to limitations, recall bias could have impacted the participant's ability to discuss their experience with the intervention. It could be beneficial to interview participants closer to the completion of their intervention time or in accordance with user design theory, which may suggest interviewing in parallel with the intervention.

Conclusions

We found that Horyzon's therapeutic social network fostered connection and understanding between young people. It also aided in the creation of an embodied experience that afforded young people with FEP a sense of self-recognition and belonging over the longer-term. However, although we found that most young people had strong positive experiences of social connection on Horyzons, we also found that they experienced significant barriers that could substantively interrupt their ability to use the platform. We found that social anxiety, paranoia, internalized stigma, lack of autonomy, and social protocol confusion interfered with young people's usage of the platform. From a design perspective, digital interventions are flexible and thus equipped to begin addressing these implications by providing customizable and personalized treatment options that account for varying levels of social connection and psychological need that could otherwise interrupt young people's usage of social media–based interventions.

Acknowledgments

The authors would like to thank all of the young people who participated in this study and shared their time, experiences, and thoughts with us. LV was supported by the Mental Illness Research Fund from the State Government of Victoria. CM was supported by an Australian Government Research Training Program Scholarship. SO was supported by a rebound National Health and Medical Research Council (NHMRC) project grant. MA-J was supported by an NHMRC investigator grant APP1177235. SB was supported by a McCusker Charitable Foundation fellowship.

Authors' Contributions

LV conducted participant interviews and data analysis. The analysis was supervised by SB. The study was cosupervised by MA and SB. All authors have reviewed the final paper.

Conflicts of Interest

None declared

References

1. da Rocha BM, Rhodes S, Vasilopoulou E, Hutton P. Loneliness in psychosis: a meta-analytical review. Schizophr Bull 2018 Jan 13;44(1):114-125 [FREE Full text] [doi: 10.1093/schbul/sbx036] [Medline: 28369646]
2. Santesteban-Echarri O, Paino M, Rice S, González-Blanch C, McGorry P, Gleeson J, et al. Predictors of functional recovery in first-episode psychosis: a systematic review and meta-analysis of longitudinal studies. Clin Psychol Rev 2017 Dec;58:59-75. [doi: 10.1016/j.cpr.2017.09.007] [Medline: 29042139]
3. Lim MH, Gleeson JF, Alvarez-Jimenez M, Penn DL. Loneliness in psychosis: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2018 Mar;53(3):221-238. [doi: 10.1007/s00127-018-1482-5] [Medline: 29327166]
4. Cotton SM, Lambert M, Schimmelmann BG, Filia K, Rayner V, Hides L, et al. Predictors of functional status at service entry and discharge among young people with first episode psychosis. Soc Psychiatry Psychiatr Epidemiol 2017 May;52(5):575-585. [doi: 10.1007/s00127-017-1358-0] [Medline: 28233045]
5. Sündermann O, Onwumere J, Kane F, Morgan C, Kuipers E. Social networks and support in first-episode psychosis: exploring the role of loneliness and anxiety. Soc Psychiatry Psychiatr Epidemiol 2014 Mar;49(3):359-366 [FREE Full text] [doi: 10.1007/s00127-013-0754-3] [Medline: 23955376]
6. Lim MH, Gleeson JF, Rodebaugh TL, Eres R, Long KM, Casey K, et al. A pilot digital intervention targeting loneliness in young people with psychosis. Soc Psychiatry Psychiatr Epidemiol 2019 Mar 14:- epub ahead of print. [doi: 10.1007/s00127-019-01681-2] [Medline: 30874828]
7. Alvarez-Jimenez M, Gleeson J, Bendall S, Penn D, Yung A, Ryan R, et al. Enhancing social functioning in young people at ultra high risk (UHR) for psychosis: a pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
8. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104 [FREE Full text] [doi: 10.1136/bmjopen-2018-024104] [Medline: 30782893]
9. Alvarez-Jimenez M, Alcazar-Corcoles M, González-Blanch C, Bendall S, McGorry P, Gleeson J. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
10. Lal S, Nguyen V, Theriault J. Seeking mental health information and support online: experiences and perspectives of young people receiving treatment for first-episode psychosis. Early Interv Psychiatry 2018 Jun;12(3):324-330. [doi: 10.1111/eip.12317] [Medline: 26810026]
11. McEnery C, Lim MH, Knowles A, Rice S, Gleeson J, Howell S, et al. Development of a moderated online intervention to treat social anxiety in first-episode psychosis. Front Psychiatry 2019;10:581 [FREE Full text] [doi: 10.3389/fpsyt.2019.00581] [Medline: 31474889]
12. Rice S, Robinson J, Bendall S, Hetrick S, Cox G, Bailey E, et al. Online and social media suicide prevention interventions for young people: a focus on implementation and moderation. J Can Acad Child Adolesc Psychiatry 2016;25(2):80-86 [FREE Full text] [Medline: 27274743]
13. Valentine L, McEnery C, D'Alfonso S, Phillips J, Bailey E, Alvarez-Jimenez M. Harnessing the potential of social media to develop the next generation of digital health treatments in youth mental health. Curr Treat Options Psych 2019 Sep 11;6(4):325-336. [doi: 10.1007/s40501-019-00184-w]
14. Nowland R, Necka EA, Cacioppo JT. Loneliness and social internet use: pathways to reconnection in a digital world? Perspect Psychol Sci 2018 Jan;13(1):70-87. [doi: 10.1177/1745691617713052] [Medline: 28937910]
15. Maher CA, Lewis LK, Ferrar K, Marshall S, de Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
16. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
17. D'Alfonso S, Phillips J, Valentine L, Gleeson J, Alvarez-Jimenez M. Moderated online social therapy: viewpoint on the ethics and design principles of a web-based therapy system. JMIR Ment Health 2019 Dec 4;6(12):e14866 [FREE Full text] [doi: 10.2196/14866] [Medline: 31799937]
18. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to ehealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
19. Merriam SB. Qualitative Research: A Guide to Design and Implementation. New York, USA: John Wiley & Sons; 2015.

20. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
21. Smith JA, Osborne M. Interpretative phenomenological analysis. In: Breakwell G, editor. Doing Social Psychology Research. Chicester, UK: Wiley-Blackwell; 2004:229-254.
22. McGaha KK, D'Urso PA. A non-traditional validation tool: using cultural domain analysis for interpretive phenomenology. Int J Soc Res Methodol 2019 May 27;22(6):585-598. [doi: 10.1080/13645579.2019.1621474]
23. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015 Sep;25(9):1212-1222. [doi: 10.1177/1049732315588501] [Medline: 26184336]
24. Leggatt M, Woodhead G. Family peer support work in an early intervention youth mental health service. Early Interv Psychiatry 2016 Oct;10(5):446-451. [doi: 10.1111/eip.12257] [Medline: 26213221]
25. Kilpatrick E, Keeney S, McCauley C. Tokenistic or genuinely effective? Exploring the views of voluntary sector staff regarding the emerging peer support worker role in mental health. J Psychiatr Ment Health Nurs 2017 Sep;24(7):503-512. [doi: 10.1111/jpm.12391] [Medline: 28423216]
26. Deci E, Ryan R. Self-determination theory. In: Wright JD, editor. International Encyclopedia of the Social & Behavioral Sciences. London, UK: Elsevier; 2015:486-491.
27. Smith W, Ploderer B, Wadley G, Webber S, Borland R. Trajectories of Engagement and Disengagement with a Story-Based Smoking Cessation App. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI'17; May 6-11, 2017; Denver, USA p. 3045-3056. [doi: 10.1145/3025453.3026054]
28. Schueller SM, Stiles-Shields C, Yarosh L. Online treatment and virtual therapists in child and adolescent psychiatry. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):1-12 [FREE Full text] [doi: 10.1016/j.chc.2016.07.011] [Medline: 27837935]

Abbreviations

FEP: first-episode psychosis
IPA: interpretative phenomenological analysis
MOST: moderated online social therapy
NHMRC: National Health and Medical Research Council
RCT: randomized controlled trial
SNS: social network site

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by P Navarro, K Fauvelle; comments to author 29.01.20; revised version received 17.02.20; accepted 21.02.20; published 26.06.20.

Please cite as:
Valentine L, McEnery C, O'Sullivan S, Gleeson J, Bendall S, Alvarez-Jimenez M
Young People's Experience of a Long-Term Social Media–Based Intervention for First-Episode Psychosis: Qualitative Analysis
J Med Internet Res 2020;22(6):e17570
URL: https://www.jmir.org/2020/6/e17570
doi:10.2196/17570
PMID:32384056

Original Paper

Evaluating Engagement in a Digital and Dietetic Intervention Promoting Healthy Weight Gain in Pregnancy: Mixed Methods Study

Jane C Willcox[1], PhD; Daniel Chai[2]; Lawrence J Beilin[3], MD; Susan L Prescott[3], MD, PhD; Desiree Silva[4], MD, PhD; Cliff Neppe[4], MD; Rae-Chi Huang[2], MD, PhD

[1]La Trobe University, Bundoora, Australia
[2]Telethon Kids Institute, Perth, Australia
[3]The University of Western Australia, Perth, Australia
[4]Joondalup Health Campus, Shenton Avenue, Perth, Australia

Corresponding Author:
Rae-Chi Huang, MD, PhD
Telethon Kids Institute
Hospital Avenue
Nedlands
Perth, 6009
Australia
Phone: 61 0402131555
Email: rae-chi.huang@telethonkids.org.au

Abstract

Background: Early excess and inadequate gestational weight gain (GWG) have been associated with negative outcomes for mother and child. The use of digital media to deliver pregnancy lifestyle interventions is increasing, but there is little data on participant engagement. The Pregnancy Lifestyle Activity and Nutrition (PLAN) intervention pilot study was an electronic health and dietetic-delivered intervention program promoting healthy GWG in early pregnancy.

Objective: This study aims to explore the interactions of participants with the program and to assess its acceptability.

Methods: This study uses both quantitative and qualitative methods using data from parent randomized controlled trial (ACTRN12617000725369). Quantitative data from 22 participants in the intervention arm who completed the study provided measures of the interactions participants had with the digital components of the program and with dietetic consultations. A descriptive qualitative analysis employed semistructured interviews with 9 participants to elicit views on the acceptability of the intervention and its components.

Results: The electronic delivery of information and recording of weight from 8 to 20 weeks of gestation were universally accepted. Component (face-to-face dietitian, weight tracker, website information delivery, and SMS goal prompting) acceptability and engagement differed between individuals. A total of 4 key themes emerged from the qualitative analysis: supporting lifestyle change, component acceptability and value, delivery platforms, and engagement barriers.

Conclusions: The PLAN intervention and its delivery via a blend of personal dietetic consultations and digital program delivery was found to be acceptable and valuable to pregnant women. Individuals responded differently to various components, emphasizing the importance of including women in the development of lifestyle interventions and allowing participants to choose and tailor programs. Larger randomized controlled trials using these insights in a broader section of the community are needed to inform the iterative development of practical, time-efficient, and cost-effective ways of supporting optimal GWG with the potential to optimize outcomes for pregnant women and their child.

(J Med Internet Res 2020;22(6):e17845) doi:10.2196/17845

KEYWORDS

pregnancy; internet-based intervention; patient participation; qualitative research; eHealth; body weight; obesity

Introduction

Background

Gestational weight gain (GWG) has emerged as an important health variable contributing to the health of both mother and child. Weight gained outside the recommended Institute of Medicine guidelines [1] across all prepregnancy BMIs is associated with negative short- and long-term health outcomes [2,3]. For example, excess GWG increases the risk of antenatal hypertensive disorders, gestational diabetes, and atypical delivery outcomes [2,4-8], along with maternal, infant, and later life overweight or obesity [4,6]. Over the last two decades, women have increasingly failed to meet GWG recommendations, with more recently an estimated 30% to 65% of women in developed countries, across all prepregnancy BMI categories, exceeding the recommended guidelines for GWG [9,10]. These figures highlight the need for better education directed toward prevention.

Research interest in GWG has seen an increasing number of studies promoting healthy GWG through diet only, exercise only, and combined diet and exercise interventions, compared with standard care [11]. A limitation of many of these interventions is the reliance on support from research staff or health professionals, restricting the reach and scalability to larger sections of the community [12]. More recently, a small number of pilot studies have investigated the digital delivery of GWG interventions using internet and mobile phones (electronic health [eHealth] and mobile health [mHealth]), and the results from these studies show promise [13-15]. Digital delivery potentially offers an opportunity to augment health professional care and provide trusted source information across demographic groups through a low-cost, easy access method [16]. Although a 2017 meta-analysis of the effectiveness of eHealth technologies on GWG found a nonsignificant pooled result in only six studies, it emphasized that the engagement of participants was critical to the success of digitally delivered interventions [13].

Understanding how participants engage and the experiences they have with interventions is crucial to assist in their evaluation and to inform improvement [17,18]. According to O’Brien et al [19], *engagement* may be defined as “the ability to engage and sustain engagement in digital environments.” A number of papers have emphasized the importance of understanding and incorporating the views and engagement metrics of stakeholders into the evaluation of digital interventions [17,18]. For example, in a review of mobile technology in psychosocial and health behavior treatments by Heron and Smyth [17], the authors suggest that interventions need to be more sensitive to individual characteristics and needs of stakeholders. They argued that incorporating end-user feedback into evaluation will facilitate the feasibility and acceptability of interventions. There is a scarcity of literature evaluating participants' engagement with pregnancy and GWG interventions.

The Pregnancy Lifestyle Activity and Nutrition (PLAN) randomized controlled trial aimed to promote healthy GWG in early pregnancy via an eHealth and dietetic-delivered intervention [20]. Women were recruited between 8 and 11 weeks of gestation and randomized to the intervention or routine antenatal care. The 12-week intervention involved one dietetic one-on-one consultation, examining diet and physical activity, and a web-based and SMS program providing diet, physical activity, and well-being advice. The digital component of the program included tracking of GWG, lifestyle goal setting and feedback and diet, physical activity, and mental health information delivered weekly. Participants were then followed up to 3 months postpartum.

A total of 57 women, with a mean age of 33.3 (SD 2) years and a mean gestational age of 9.2 (SD 1.2) weeks, were recruited. The mean BMIs of women in the control and intervention groups at enrollment were 25.3 (SD 5.3) kg/m^2 and 26.0 (SD 1.3) kg/m^2, respectively. A total of 43/57 (75%) participants (control group: 19 and intervention group: 24) completed the study. Of the 24 participants in the intervention arm, 22 completed the 12-week intervention program. Although the pilot-sized numbers precluded the power to detect differences in GWG, as well as pregnancy and infant outcomes, infants weighed less in the intervention group (6193 g versus 5405 g; P=.01) at 3 months. Furthermore, participants in the intervention arm demonstrated a mean decrease in total fat consumption, increase in fruit serves per day, and score ranking perceived importance of dietary change, compared with participants in the control group. Although limited in number, these pilot results and learnings contribute to the literature on dietetic and digitally delivered antenatal interventions.

Objective

Consistent with the postevaluation phase of the development and evaluation processes for digitally delivered interventions [18], this study aimed to explore women’s engagement and perspectives on the PLAN eHealth and dietetic intervention [20].

Methods

This mixed methods study used quantitative assessment for the interaction with the PLAN intervention and qualitative assessment to investigate women’s engagement with and acceptance of the PLAN intervention.

The mHealth Development and Evaluation framework [18] and the Process-Evaluation Plan for Assessing Health Programme Implementation [21] informed the development of this study. The Consolidated Reporting Criteria for Qualitative Studies [22] informed the reporting of this study (Multimedia Appendix 1).

Ethics approval was obtained from the human research ethics committees of Joondalup Health Campus (October 17, 2015; ethics number 1525) and St John of God Health Care (May 26, 2016; ethics number 873). The trial was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12617000725369).

Quantitative Assessment of Intervention Engagement

Intervention interaction was measured post hoc by program-generated data. The data were extracted from participants in the intervention arm who participated in the 12-week intervention (n=22).

Data Collection

The characteristics of the participants, including sociodemographic and GWG, were collected from the program data. Indicators of interaction were extracted from the program data files and included retention (numbers completing the trial), dose delivered (contacts delivered), and intervention engagement (contact with the program elements).

The contacts delivered included dietetic consultation, emails with links to lifestyle information and weight tracking prompting, and SMS for goal prompting. Contacts with the program elements included attendance at the dietetic consultation, accessing the web-based app, input into weight tracking, and replies to the goal accomplishments. Data on the frequency and depth of website access were not available.

Data Analysis

Quantitative interaction and participants' characteristics were analyzed with categorical variables reported as numbers and percentages and continuous variables reported as means and SDs.

Qualitative Assessment of Intervention Engagement and Acceptance

A qualitative descriptive research methodology [23,24] using semistructured interviews was employed to obtain in-depth data from consenting women.

Study Participants

Women who had completed the active 12-week phase of the PLAN intervention (n=22) were invited to participate in the qualitative review by the PLAN research coordinator.

Data Collection

Telephone interviews were conducted by an independent female interviewer trained in qualitative research methods and not involved in the PLAN study (JCW, a research fellow). The participants were informed that JCW was an independent researcher from a different university and had experience in this qualitative area of research. The content of the standard interview guide (Multimedia Appendix 2) was developed by R-CH. The semistructured and structured questions aimed to elicit women's views on the PLAN intervention, its components, and potential improvements. In addition, women were asked to rate the experience of the PLAN intervention on helping follow a healthier lifestyle during pregnancy (1=not good and 10=very good). The interviews were digitally recorded with the consent of the participants. Field notes were collected by JCW during and after the telephone interviews.

Data Analysis

Data immersion, coding, category creation, and thematic analysis were used to find repeated patterns of meaning across datasets [25,26]. The researchers (JCW and R-CH) used an inductive approach using raw data to derive themes through interpretations made from the raw data [27].

Results

Participants

The sociodemographic and GWG characteristics of the women in the total PLAN study are reported elsewhere [20]. The sociodemographic and GWG characteristics of the 22 participants in this substudy are provided in Table 1. The mean age of participants was 33.0 (SD 4.1) years, with a mean gestational age of 9.0 (SD 1.2) weeks at recruitment. The mean self-reported prepregnancy BMI was 25.4 (SD 4.7) kg/m^2, and one-fifth of the participants (21%) gained weight within the GWG guidelines. The majority of the participants (91%) had private antenatal care, and nearly two-thirds (64%) of those reported that their education held postsecondary qualifications.

Table 1. Characteristics of participants in the active Pregnancy Lifestyle Activity and Nutrition intervention arm and qualitative interviews.

Participant characteristic	Participants in the intervention arm who completed the study (n=22)	Qualitative interviews (n=9)
Maternal age (years), mean (SD)	33.0 (4.1)	33.0 (3.2)
Enrollment gestational age (weeks), mean (SD)	9.0 (1.2)	9.0 (1.3)
Reported prepregnancy BMI (kg/m^2), mean (SD)	25.4 (4.7)	26.7 (4.3)
Measured enrollment BMI (kg/m^2), mean (SD)	26.2 (4.9)	28.3 (4.5)
Enrollment BMI category (kg/m^2), n (%)		
Underweight (<18.5)	0 (0)	0 (0)
Normal weight (18.5-24.9)	10 (45)	3 (33)
Overweight (25.0-29.9)	8 (36)	3 (33)
Obese (≥30.0)	4 (18)	3 (33)
Ethnicity, n (%)		
White	22 (100)	9 (100)
Private health care, n (%)	20 (91)	9 (100)
Household income, n (%)		
<AUD $100,000 (US $64,368)	2 (9)	0 (0)
AUD $100,001-AUD $150,000 (US $64,368-US $96,552)	6 (27)	1 (11)
>AUD $150,000 (US $96,552)	12 (54)	5 (56)
Did not answer	2 (9)	3 (33)
Highest level of education, n (%)		
Tertiary	12 (54)	5 (56)
Other (diploma or trade)	2 (9)	0 (0)
Secondary or below	5 (23)	1 (11)
Unknown	3 (14)	3 (33)
Parity, n (%)		
Nulliparous	9 (41)	3 (33)
Multiparous	12 (54)	3 (33)
Unknown	1 (4)	3 (33)
Total gestational weight gain (kg), mean (SD)	13.1 (3)	14.7 (4)
Gestational weight gain within Institute of Medicine guidelines[a], n (%)		
Below	5 (21)	1 (11)
Within	5 (21)	2 (22)
Exceed	12 (54)	6 (67)

[a] Intervention period.

Quantitative Assessment of Intervention Interactions: Intervention Interaction Data

All women accessed the website during the course of the program. Nearly all women (21/22, 96%) attended the dietetic visit and completed the accelerometer tracking (Table 2). Over the 12-week active intervention period, 31 emails were sent to each participant to notify them of a new release of nutrition, physical activity, and GWG information related to their pregnancy stage. In addition, over this period, a mean of 4.46 (SD 6.24) emails were sent to remind participants to input their weight on the website if this had not been completed. The total number of inputs for weight totaled 23.17 (SD 24.21), ranging from 1 to 92. One-fourth of participants (n=5/22, 23%) received emails notifying them of excessive weight gain (3 participants received 1 notification, 1 participant received 2 notifications, and 1 participant received 3 notifications).

Table 2. Intervention interaction data (n=22).

Components of the Intervention	Values	Range
Program component delivery, n (%)		
Dietetic visit with personalized nutrition and physical activity assessment and education	21 (96)	N/A[a]
Accelerometer wear and feedback	21 (96)	N/A
Website access[b]	22 (100)	N/A
Email engagement[c], mean (SD)		
Emails sent regarding new material on the web	31 (0)	31-31
Number of emails received reminding to input weight	4.46 (6.24)	0-23
Number of emails received warning of excess gestational weight gain	0.33 (0.75)	0-3
SMS messages sent prompting SMART[d] goal	8.5 (1.5)[e]	3-9
Reply to SMS message prompting SMART goal setting	2.96 (2.65)	0-8
Weight input: active self-monitoring	N/A	N/A
Number of weight inputs by participants	23.17 (24.21)	1-92

[a] Not applicable.

[b] Website access data only. Frequency of access not available.

[c] Value reflects the mean number of participants (SD).

[d] SMART: specific, measurable, attainable, realistic, and time-based.

[e] A total of 19 participants received all 9 reminders over the course of the active intervention, 3 opted out at weeks 4, 6, and 9, and hence, they received 3, 5, and 8 SMS messages, respectively.

Women were asked to set a specific, measurable, attainable, realistic, and time-based (SMART) goal [28] to assist them with lifestyle behavior change. SMS messages were sent prompting engagement with the SMART goals women set at the program commencement, with an invitation to reply via return SMS. Overall, 3 women opted out of receiving these messages partway through the intervention, receiving between 3 and 8 SMS messages prompting SMART goal engagement. The remaining participants received the full complement of 9 SMS messages prompting SMART goal engagement over the active 12-week period. Over the 12-week period, women replied on average 2.96 (SD 2.65, range 0-8) times to these messages, with 77% (n=17/22) of participants responding at least once with a range of 0 to 8. The SMS replies from the participants indicating if the SMART goal had been attained was positive on 86%, negative on 10%, and neutral on 4% of occasions.

Qualitative Assessment of the Intervention: Study Participants' Characteristics

In total, 11 women agreed to participate in the qualitative study (n=11/22, 50%), with 2 participants withdrawing before the interview. Of the 11 women, 9 were consented and interviewed, with 4 women having completed the total study, including the active and follow-up phases (range 38 weeks' gestation to 16 weeks postpartum). Of these 5 completed the active phase (range 22-28 weeks' gestation) and were yet to complete their pregnancy.

The mean age of women was 33.0 (SD 3.2) years, and enrollment BMI was 28.3 (SD 4.5) kg/m^2. One-third of women were of healthy weight, one-third were overweight, and one-third were obese. All women were identified as white, and 43% of the women had completed tertiary education. The interview length ranged from 22 to 36 min.

Emergent Themes

All participants spoke about their experiences with the PLAN intervention and the acceptability and usefulness of the intervention and its components and offered suggestions for improvement to support them in attaining a healthy GWG and pregnancy lifestyle. Saturation of themes appeared after 7 participants. In all, 4 key areas of focus for the women included the effectiveness of the PLAN intervention to encourage lifestyle change, individual component acceptability, delivery platforms, and barriers to engagement with the PLAN intervention.

Pregnancy Lifestyle Activity and Nutrition Intervention Supporting Lifestyle Change

In general, women reported a positive experience with the PLAN intervention. They were asked to rate their experience of the PLAN intervention with helping follow a healthier lifestyle during pregnancy with *one* (the lowest rating) and *ten* (the highest rating). A total of 2 participants rated the program a *ten*, 3 rated it an *eight*, 3 rated it a *seven*, and 1 rated it a *five*. A participant, who rated the program an *8*, qualified her answer by rating the active part of the program an *8* and by rating the nonactive part after 20 weeks a *two*.

When asked if they would recommend the program to a friend, all but 1 woman answered *yes* and 1 answered *maybe*. Again, 1 participant qualified the answer by saying *yes* if the program was fully active and *no* if it only involved data collection.

The most valuable aspect of PLAN cited by many women was the frequent contact and the level of support and encouragement provided. Overall, women conveyed valuing and liking “the extra support in the background”:

> *It was nice to feel supported and feel like you are on track. I really appreciated the ladies that worked with the study.* [Participant 9]
>
> *I felt cared for and nurtured through this [pregnancy] period.* [Participant 7]

Different Program Components Judged Acceptable and Valuable

Participants discussed the program components as mutually exclusive elements with different participants citing components that they found of value.

Dietitian Consultation

The dietetic consultation, which included assessment of food records, was highly valued by the majority of participants. A total of 4 participants expressed positive *surprise* with the extent of information and assistance that they received, particularly around the food groups that they were missing in their diets.

Another 2 participants suffered *morning sickness* and found the dietitian consultation ill-timed for their medical state. They reported finding it difficult to receive information that encouraged exercise and eating well and would have welcomed a later appointment or advice related to *morning sickness*. They expressed a desire to follow up on the recommended dietary changes or have a reassessment once their *morning sickness* had dissipated.

When asked, all women conveyed that they would find it helpful to have another consultation at around 16 to 20 weeks of gestation:

> *It would be great to get it again at 16-20 weeks. It is good to get the feedback.* [Participant 8]

Weight Tracker

Most women reported using the weight tracker at least once per week. Many women found value in prompting to weigh, recording their weight, and seeing a visual representation of the weight trajectory. A total of 4 women nominated the weight tracker as one of the most valuable aspects because of its help in “seeing how it progressed” and “reinforced the importance of eating well and being active”:

> *I looked forward to doing it. It was good to be prompted.* [Participant 7]
>
> *It was reassuring as a woman.* [Participant 9]

All but 2 women stated that they received reminders to weigh (once per week); 4 women described the frequency of weight input and reminder as acceptable, 1 as too infrequent, and 2 as too frequent.

In all, 3 women did not recall receiving feedback or comments from the weight tracker. One woman reported finding positive feedback but further commented that her weight stayed within the recommended quadrant. Moreover, 3 other women conveyed that they were not happy with the negative feedback, which they found “unhelpful.” They desired only constructive feedback and more of a sense of what to alter:

> *I would have liked more mindful about your lifestyle feedback.* [Participant 7]

Emails, Websites, and Health Information

All but 2 women reported receiving regular notifications for the release of nutrition, physical activity, and GWG information on the website/web-based app related to their pregnancy stage.

A total of 4 women reported reading all the information, 2 reading more than two-thirds, and 3 had not read any information. Many of the women who read the material found that it provided positive information for a lifestyle change and reinforced key health messages. One woman who read the information found the information did not meet her needs by being “too light”:

> *I liked the simple terms, not too complex. It made you think about your lifestyle.* [Participant 6]
>
> *[It] reinforced the importance of eating well and being active.* [Participant 1]

Although many participants found the quantity of information appropriate, 1 woman found the quantity of information overwhelming. This was partly because of not being able to exercise for medical reasons and found the information that did not relate to her difficulty:

> *It got a bit overwhelming. I felt bad that I couldn’t read and do things.* [Participant 8]

Most of the women who read the information indicated that they liked the information being “fed slowly.” The quantity of nutrition and exercise information each release was deemed “about right.”

All women expressed a desire for the information to continue beyond 20 weeks until pregnancy completion with information tailored to the issues relevant to pregnancy at that stage (eg, reflux). There were 2 women that stated preferences for the information to be provided at the start of the intervention, allowing them to access the correct information when required.

Participants suggested a range of opportunities for improvements to the information, with the most frequently suggested being more practical information about healthy eating, including meal plans and recipes. Additional structured guidance about physical activity was desired by 3 women with more information about the later stages of pregnancy. A few women desired more well-being information:

> *More information on well-being I think. I was neglectful of myself.* [Participant 7]

Moreover, 2 women suggested a more up-to-date “look and feel” and cited “Mind the Bump” and “Huggies” websites as having looks and functionality that they liked.

Goal Setting

Setting SMART goals was described as being “helpful” by some participants. The goals stopped them from becoming “complacent” and were good “external motivation.” The importance of the timing of text messages was mentioned by a few:

> *It was good to have a reminder.* [Participant 8]

> *The texts were frustrating when not in that space.* [Participant 4]

Another woman reported that she set a physical activity goal that she had already met for the sake of the program. She conveyed that she was not a "big goal setter."

Short-term goals that changed regularly were suggested by 4 women, and 2 women would find continuation of the goal setting beyond the active 12-week program.

Delivery Platform as a Positive and Familiar Source of Information and Contact

The electronic delivery of information and recording of weight was seen as positive, with women being familiar with this mode of delivery:

> *We are used to those sort of tools with Fitness Pals and others.* [Participant 9]

When prompted, women suggested multiple delivery platforms to meet different needs for different times:

> *Everyone loves an app now.* [Participant 9]
>
> *Texts with links to the information would be good as well as an app and a website.* [Participant 7]

Some women viewed that the addition of a private Facebook page or community forum would be positive, with its ability to allow interaction between women of different pregnancy stages:

> *It would be great to build rapport with other ladies going through the pregnancy journey.* [Participant 7]

Others were concerned about the intrusiveness of Facebook and not wanting to engage in a community where there was a burden of stress:

> *I have avoided it [Facebook] because of all the worriers on there. It depends on how it was structured.* [Participant 9]

Barriers to Engaging in a Pregnancy Lifestyle Program

Barriers to engaging with the PLAN program were mentioned by all participants in the context of barriers to program engagement and barriers to changing lifestyles.

Program-related difficulties included "glitches," "freezing," or "incorrect plotting" being reported by many participants. One participant recounted receiving a message telling her that she had gained excess weight when it did not plot that way on the graph. Another described weight being plotted on the graph in the incorrect section. Several women reported that the information was not accessible on mobile devices, which limited information access at convenient times.

Many women reported surprise in the cessation of the program, including the ability to track weight, beyond the active stage of the PLAN intervention at 20 weeks:

> *I was not mentally prepared for it [the weight tracker and program] to stop.* [Participant 6]

All but 1 woman wanted to be able to track weight and have the program continue till pregnancy completion to serve as a reminder to keep them on track with regard to lifestyle:

> *I would have liked to see it all the way through to after the baby was here. The numbers don't lie.* [Participant 7]

Women frequently cited barriers to maintaining a healthy lifestyle and following the guidelines set by the PLAN intervention. Morning sickness was a key reason given by 4 women, reflecting the early gestation commencement of the PLAN intervention. Work, family, and time constraints were the next most commonly discussed constraints:

> *I didn't use it all cause of working 52 hours a week.* [Participant 3]

Back pain and other pregnancy-related physical ailments were also cited, particularly as barriers to physical activity:

> *I had more pain in my second pregnancy.* [Participant 9]

A common barrier and concern expressed was the perceived inability to access ”pregnancy-safe” food when eating out. Therefore, some women concluded that it felt easier and safer to purchase unhealthy food that was more likely to be food safe and listeria free.

Discussion

This paper investigated both the quantitative and qualitative engagement of women in the PLAN intervention. The data contribute to the literature on the engagement and acceptability of digitally supported interventions in pregnancy. Of the women who completed the program, the majority accessed the dietetic consultation and the website. Most women read the health information provided and interacted and tracked GWG, whereas fewer women interacted with or replied to the SMART goal–monitoring SMS messages. A total of 4 key areas of focus were elucidated from the qualitative assessment, including the effectiveness of the PLAN intervention to encourage lifestyle change, individual component acceptability, delivery platform acceptability, and barriers to engagement with the PLAN intervention. Overall acceptance of the program was high, with the social support being offered identified as key. The components (dietitian consultation, weight tracker, health information, and goal setting) were identified as mutually exclusive elements of the intervention with variable acceptance. Although there was a consensus that the electronic delivery of information and recording of weight was positive and universally accepted, the cited barriers to adherence varied among the women. Suggestions for improvement of trial design, including an extension of the active program throughout pregnancy, will be incorporated into future program iterations.

The perceived support provided by the intervention program was a strong theme elicited from the women. This supports the findings of others who have found that health interventions, including digital support, can engender a continued sense of support and community [29-31]. Indeed, social support is an important technique [32] in lifestyle pregnancy interventions that have successfully altered GWG [15,21,33]. Digitally delivered interventions provide the opportunity for more frequent interactions delivered beyond the traditionally delivered environments [31].

This intervention mapped many of the key behavioral change variables required for behavior change [34], including providing information on the behavior-health link, goal setting, review of goals, and self-monitoring. The inclusion of behavior change techniques may be most effective when combined with dietary advice in GWG interventions [35]. Digital technologies capitalize on the ability to extend the reach of intervention delivery beyond a health professional appointment and expand the range of behavior change techniques that can be included [36]. Evidence for blended health professional and digitally delivered interventions is still limited to a small number of studies, and further research is needed to explore what types of blended interventions may be most effective and acceptable to both women and health professionals.

Individual participants identified greater affinity for different intervention components. Although some women identified the weight tracker as one of the most valuable components, others nominated the health information or dietitian visit as key. This highlights the importance of offering a suite of products and interactions that allows women to tailor their own programs and interactions according to their preferences, including personal physical and mental health and social support. This is consistent with a study by Willcox et al [37], where both health professionals and women articulated that the individual requirements of women would be best served by interventions that integrated multiple technology elements. This was seen to serve the individual acceptability and use of different technologies and also the needs of different learning styles. Women saw their ability to access different technologies and elements as a way to self-manage or control information acquisition that was unavailable in traditional care models and information sources. Heron and Smyth argue in their review of mobile technology in psychosocial and health behavior research that interventions need to be more sensitive to the individual characteristics and needs of stakeholders [17]. Further investigation with larger trials will be required to understand the underlying differences in characteristics that underpin this and understand how researchers may segment users and target them most appropriately.

One important finding of this study is that the recording of and education around weight was universally accepted by women. The weight tracker was consistently used by many women, and the qualitative interviews elicited positive views on the practice and prompting of weighing. Self-monitoring is known to be a powerful behavior change technique and is frequently adopted in technology interventions [38]. Feedback from all but 1 woman demonstrated the desire to be able to track weight though to pregnancy completion. Previous research has highlighted the need for women to understand their targeted GWG and allow tracking [39]. Although GWG is well understood to impact the health outcomes of both mother and child, concerns are often raised by health professionals as to the appropriateness of weight discussions with pregnant women [40,41]. Health professionals often cite concern for the physical and psychological health of pregnant women and worries about the perceived negative impacts of weight discussions. This study adds to the data supporting that the majority of women are accepting of GWG discussions and tracking. Further research could engage a wide range of health professionals in supporting digital health augmentation to their counseling to ensure that GWG messages are consistent and professionals are equipped with the necessary skills to address weight management in pregnancy.

Strengths and Limitations

A strength of this study was the use of a mixed methods design incorporating both quantitative engagement data and qualitative engagement and acceptability data. The pilot size of the original PLAN study, and hence the qualitative work, may be viewed as a limitation. Despite this, the application of qualitative research methods was able to define certain themes. Furthermore, the additional corroborating objective measures of engagement, such as frequency of weight entry and SMS response, correlated well with what was reported in the interviews, providing greater evidence for the integrity of the information obtained. Larger-scale studies with analysis of participant engagement and health outcomes will allow further exploration of engagement variables. Another limitation of the study was the recruitment of predominantly private patients from an upper socioeconomic stratum, with the complexity of recruiting early gestational age participants under public care. Recruiting more broadly and tailoring and testing the intervention on a broader section of the community is required. That only half of the women agreed to participate in the qualitative interviews may be viewed as a limitation. Although saturation was reached after 7 participants, the nonresponders may have held alternate views.

Conclusions

This study investigated the engagement with and acceptability of a combination of eHealth and dietetic intervention to promote healthy GWG in pregnant women. Women universally found the blended model of a dietitian consult and digital information and tracking acceptable and opportune. However, participants responded more favorably to different elements of the intervention and described a range of factors to enhance future interventions in preventing excess weight gain in pregnancy. This pilot study contributes to the literature on participant engagement and provides participant-related themes, including the importance of choice and tailoring, to inform future research.

Acknowledgments

JCW was supported by a La Trobe University Research Fellowship. R-CH was supported by the National Health and Medical Research Council Fellowship GNT1053384. This study was supported by the ORIGINS Project, Joondalup Health Campus. The study was also supported by a seeding grant from the Telethon Kids Institute.

Conflicts of Interest

SLP reports personal fees from Swiss Biostime, Bayer Probiotic Advisory Board, Danone, and Sanofi Advisory Group outside the submitted work and research grants from the National Health and Medical Research Council (NHMRC) of Australia and the Paul Ramsay Foundation.

Multimedia Appendix 1
The Consolidated Reporting Criteria for Qualitative Studies for this study.
[PDF File (Adobe PDF File), 497 KB - jmir_v22i6e17845_app1.pdf]

Multimedia Appendix 2
Interview questions (PLAN intervention).
[DOCX File , 15 KB - jmir_v22i6e17845_app2.docx]

References

1. National Research Council, Institute of Medicine, Families Board on Children, Youth, Food and Nutrition Board, Committee to Reexamine IOM Pregnancy Weight Guidelines, Rasmussen K, et al. Weight Gain During Pregnancy: Reexamining the Guidelines. Washington, DC: National Academies Press; 2009.
2. Viswanathan M, Siega-Riz A, Moos MK, Deierlein A, Mumford S, Knaack J, et al. Outcomes of maternal weight gain. Evid Rep Technol Assess (Full Rep) 2008 May(168):1-223. [Medline: 18620471]
3. Cedergren M. Effects of gestational weight gain and body mass index on obstetric outcome in Sweden. Int J Gynaecol Obstet 2006 Jun;93(3):269-274. [doi: 10.1016/j.ijgo.2006.03.002] [Medline: 16626716]
4. Kieffer EC, Tabaei BP, Carman WJ, Nolan GH, Guzman JR, Herman WH. The influence of maternal weight and glucose tolerance on infant birthweight in Latino mother-infant pairs. Am J Public Health 2006 Dec;96(12):2201-2208. [doi: 10.2105/AJPH.2005.065953] [Medline: 17077395]
5. Kiel DW, Dodson EA, Artal R, Boehmer TK, Leet TL. Gestational weight gain and pregnancy outcomes in obese women: how much is enough? Obstet Gynecol 2007 Oct;110(4):752-758. [doi: 10.1097/01.AOG.0000278819.17190.87] [Medline: 17906005]
6. Jensen DM, Ovesen P, Beck-Nielsen H, Mølsted-Pedersen L, Sørensen B, Vinter C, et al. Gestational weight gain and pregnancy outcomes in 481 obese glucose-tolerant women. Diabetes Care 2005 Sep;28(9):2118-2122. [doi: 10.2337/diacare.28.9.2118] [Medline: 16123476]
7. Nohr E, Vaeth M, Baker J, Sørensen TI, Olsen J, Rasmussen K. Combined associations of prepregnancy body mass index and gestational weight gain with the outcome of pregnancy. Am J Clin Nutr 2008 Jun;87(6):1750-1759. [doi: 10.1093/ajcn/87.6.1750] [Medline: 18541565]
8. Hilson J, Rasmussen K, Kjolhede C. Excessive weight gain during pregnancy is associated with earlier termination of breast-feeding among white women. J Nutr 2006 Jan;136(1):140-146. [doi: 10.1093/jn/136.1.140] [Medline: 16365073]
9. Gaillard R, Felix JF, Duijts L, Jaddoe VW. Childhood consequences of maternal obesity and excessive weight gain during pregnancy. Acta Obstet Gynecol Scand 2014 Nov;93(11):1085-1089. [doi: 10.1111/aogs.12506] [Medline: 25231923]
10. Bookari K, Yeatman H, Williamson M. Falling short of dietary guidelines-what do Australian pregnant women really know? A cross sectional study. Women Birth 2017 Feb;30(1):9-17. [doi: 10.1016/j.wombi.2016.05.010] [Medline: 27317129]
11. Muktabhant B, Lawrie T, Lumbiganon P, Laopaiboon M. Diet or exercise, or both, for preventing excessive weight gain in pregnancy. Cochrane Database Syst Rev 2015 Jun 15(6):CD007145. [doi: 10.1002/14651858.CD007145.pub3] [Medline: 26068707]
12. Yeo S, Walker JS, Caughey MC, Ferraro AM, Asafu-Adjei JK. What characteristics of nutrition and physical activity interventions are key to effectively reducing weight gain in obese or overweight pregnant women? A systematic review and meta-analysis. Obes Rev 2017 Apr;18(4):385-399. [doi: 10.1111/obr.12511] [Medline: 28177566]
13. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, et al. The effectiveness of ehealth technologies on weight management in pregnant and postpartum women: systematic review and meta-analysis. J Med Internet Res 2017 Oct 13;19(10):e337 [FREE Full text] [doi: 10.2196/jmir.8006] [Medline: 29030327]
14. Soltani H, Duxbury A, Arden M, Dearden A, Furness P, Garland C. Maternal obesity management using mobile technology: a feasibility study to evaluate a text messaging based complex intervention during pregnancy. J Obes 2015;2015:814830 [FREE Full text] [doi: 10.1155/2015/814830] [Medline: 25960889]
15. Willcox J, Wilkinson S, Lappas M, Ball K, Crawford D, McCarthy E, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728 [FREE Full text] [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
16. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]

17. Heron K, Smyth J. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
18. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mhealth interventions: examples from New Zealand. J Health Commun 2012;17(Suppl 1):11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
19. O’Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Hum-Comput Stud 2018 Apr;112:28-39 [FREE Full text] [doi: 10.1016/j.ijhcs.2018.01.004]
20. Huang R, Silva D, Beilin L, Neppe C, Mackie KE, Roffey E, et al. Feasibility of conducting an early pregnancy diet and lifestyle e-health intervention: the pregnancy lifestyle activity nutrition (PLAN) project. J Dev Orig Health Dis 2020 Feb;11(1):58-70. [doi: 10.1017/S2040174419000400] [Medline: 31391133]
21. Poston L, Bell R, Croker H, Flynn AC, Godfrey KM, Goff L, UPBEAT Trial Consortium. Effect of a behavioural intervention in obese pregnant women (the UPBEAT study): a multicentre, randomised controlled trial. Lancet Diabetes Endocrinol 2015 Oct;3(10):767-777 [FREE Full text] [doi: 10.1016/S2213-8587(15)00227-2] [Medline: 26165396]
22. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
23. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
24. Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health 2010 Feb;33(1):77-84. [doi: 10.1002/nur.20362] [Medline: 20014004]
25. Liamputtong P. Qualitative data analysis: conceptual and practical considerations. Health Promot J Austr 2009 Aug;20(2):133-139. [doi: 10.1071/he09133] [Medline: 19642962]
26. Schmidt C. The analysis of semi-structured interviews. In: Flick U, von Kardoff E, Steinke I, editors. A Companion to Qualitative Research. London, UK: Sage Publications; 2004.
27. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]
28. Bovend'Eerdt TJ, Botell RE, Wade DT. Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. Clin Rehabil 2009 Apr;23(4):352-361. [doi: 10.1177/0269215508101741] [Medline: 19237435]
29. Maloney S, Nicklen P, Rivers G, Foo J, Ooi YY, Reeves S, et al. A cost-effectiveness analysis of blended versus face-to-face delivery of evidence-based medicine to medical students. J Med Internet Res 2015 Jul 21;17(7):e182 [FREE Full text] [doi: 10.2196/jmir.4346] [Medline: 26197801]
30. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
31. Dobson R, Whittaker R, Jiang Y, Maddison R, Shepherd M, McNamara C, et al. Effectiveness of text message based, diabetes self management support programme (SMS4BG): two arm, parallel randomised controlled trial. Br Med J 2018 May 17;361:k1959 [FREE Full text] [doi: 10.1136/bmj.k1959] [Medline: 29773539]
32. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
33. Vesco KK, Karanja N, King JC, Gillman MW, Leo MC, Perrin N, et al. Efficacy of a group-based dietary intervention for limiting gestational weight gain among obese women: a randomized trial. Obesity (Silver Spring) 2014 Sep;22(9):1989-1996 [FREE Full text] [doi: 10.1002/oby.20831] [Medline: 25164259]
34. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
35. Hill B, Skouteris H, Fuller-Tyszkiewicz M. Interventions designed to limit gestational weight gain: a systematic review of theory and meta-analysis of intervention components. Obes Rev 2013 Jun;14(6):435-450. [doi: 10.1111/obr.12022] [Medline: 23534901]
36. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
37. Willcox JC, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of women and health professionals on mhealth lifestyle interventions in pregnancy: a qualitative investigation. JMIR Mhealth Uhealth 2015 Oct 28;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]
38. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]

39. Mcphie S, Skouteris H, Hill B, Hayden M. Understanding gestational weight gain: the role of weight-related expectations and knowledge. Aust N Z J Obstet Gynaecol 2015 Feb;55(1):21-26. [doi: 10.1111/ajo.12265] [Medline: 25688816]
40. Willcox JC, Campbell KJ, van der Pligt P, Hoban E, Pidd D, Wilkinson S. Excess gestational weight gain: an exploration of midwives' views and practice. BMC Pregnancy Childbirth 2012 Sep 27;12(1):102 [FREE Full text] [doi: 10.1186/1471-2393-12-102] [Medline: 23013446]
41. van der Pligt P, Campbell K, Willcox J, Opie J, Denney-Wilson E. Opportunities for primary and secondary prevention of excess gestational weight gain: general practitioners' perspectives. BMC Fam Pract 2011 Nov 4;12:124 [FREE Full text] [doi: 10.1186/1471-2296-12-124] [Medline: 22050899]

Abbreviations

eHealth: electronic health
GWG: gestational weight gain
mHealth: mobile health
PLAN: Pregnancy Lifestyle Activity and Nutrition
SMART: specific, measurable, attainable, realistic, and time-based

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by R Gaillard, L Bodnar; comments to author 27.02.20; revised version received 16.03.20; accepted 21.03.20; published 26.06.20.

Please cite as:
Willcox JC, Chai D, Beilin LJ, Prescott SL, Silva D, Neppe C, Huang RC
Evaluating Engagement in a Digital and Dietetic Intervention Promoting Healthy Weight Gain in Pregnancy: Mixed Methods Study
J Med Internet Res 2020;22(6):e17845
URL: http://www.jmir.org/2020/6/e17845/
doi:10.2196/17845
PMID:32442153

Original Paper

A Prospective Study of Usability and Workload of Electronic Medication Adherence Products by Older Adults, Caregivers, and Health Care Providers

Tejal Patel[1,2,3,4], PharmD; Jessica Ivo[1], BSc; Sadaf Faisal[1], BPharm; Aidan McDougall[2], BSc; Jillian Carducci[2], PharmD; Sarah Pritchard[2], MScOT; Feng Chang[1], PharmD

[1]School of Pharmacy, University of Waterloo, Kitchener, ON, Canada
[2]Centre for Family Medicine Family Health Team, Kitchener, ON, Canada
[3]Department of Family Medicine, DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada
[4]Schlegel-University of Waterloo Research Institute of Aging, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Tejal Patel, PharmD
School of Pharmacy
University of Waterloo
10 Victoria St S
Kitchener, ON, N2G 1C5
Canada
Phone: 1 519 888 4567 ext 21337
Fax: 1 519 888 7910
Email: t5patel@uwaterloo.ca

Abstract

Background: A decreased capacity to self-manage medications results in nonadherence, medication errors, and drug-related problems in older adults. Previous research identified 80 electronic medication adherence products available to assist patients with self-management of medications. Unfortunately, the usability and workload of these products are unknown.

Objective: This study aimed to examine the usability and workload of a sample of electronic medication adherence products.

Methods: In a prospective, mixed methods study, a sample of older adults, health care professionals, and caregivers tested the usability and workload of 21 electronic medication adherence products. Each participant tested 5 products, one at a time, after which they completed the system usability scale (SUS) and NASA-task load index (NASA-TLX), instruments that measure the usability and workload involved in using a product. Higher SUS scores indicate more user-friendliness, whereas lower NASA-TLX raw scores indicate less workload when using a product.

Results: Electronic medication adherence products required a mean of 12.7 steps (range 5-20) for the appropriate use and took, on average, 15.19 min to complete the setup tasks (range 1-56). Participants were able to complete all steps without assistance 55.3% of the time (103 out of the 186 tests were completed by 39 participants; range 0%-100%). The mean SUS and NASA-TLX raw scores were 52.8 (SD 28.7; range 0-100) and 50.0 (SD 25.7; range 4.2-99.2), respectively, revealing significant variability among the electronic medication adherence products. The most user-friendly products were found to be TimerCap travel size (mean 78.67, SD 15.57; *P*=.03) and eNNOVEA Weekly Planner with Advanced Auto Reminder (mean 78.13, SD 14.13; *P*=.049) as compared with MedReady 1700 automated medication dispenser (mean 28.63, SD 21.24). Similarly, MedReady (72.92, SD 18.69) was found to be significantly more work intensive when compared with TimerCap (29.35, SD 20.35; *P*=.03), e-pill MedGlider home medication management system (28.43, SD 20.80; *P*=.02), and eNNOVEA (28.65, SD 14.97; *P*=.03). The e-pill MedTime Station automatic pill dispenser with tipper (71.77, SD 21.98) had significantly more workload than TimerCap (*P*=.04), MedGlider (*P*=.03), and eNNOVEA (*P*=.04).

Conclusions: This study demonstrated that variability exists in the usability and workload of different electronic medication adherence products among older adults, caregivers, and clinicians. With few studies having investigated the usability and workload of electronic medication adherence products, no benchmarks exist to compare the usability and workload of these products. However, our study highlights the need to assess the usability and workload of different products marketed to assist with medication taking and provides guidance to clinicians regarding electronic medication adherence product recommendations for their patients.

Future development of electronic medication adherence products should ensure that the target populations of patients are able to use these products adequately to improve medication management.

(J Med Internet Res 2020;22(6):e18073) doi:10.2196/18073

KEYWORDS

electronic medication adherence; usability; workload; geriatrics; older adults; mobile phone

Introduction

Background

The global population is aging, and as a result, the proportion of older adults is growing rapidly. In 2017, there were 962 million individuals aged 60 years and older; by 2050, there will be 2.1 billion, accounting for 1 out of every 5 people [1]. With an increasing life expectancy, a greater proportion of the population is living well into their seventh and eighth decades, resulting in increases in the prevalence of chronic disease and comorbid medical conditions [2-4]. Indeed, in high-income countries, 86% of the burden of disease is due to noncommunicable chronic diseases; this is a phenomenon that is being replicated in middle- and low-income countries, as the ability to address infectious and parasitic diseases grows [2]. Chronic diseases such as heart disease, hypertension, diabetes, and cancer, among others, are typically managed with medications, in many cases, with multiple medications [5-7]. The use of multiple medications, commonly referred to as polypharmacy, for the treatment of comorbid conditions, brings its own set of complexities, including adverse effects, drug interactions, drug-induced disease, complex drug dosing regimens, and nonadherence, all of which increase the risk of hospitalization and mortality [5,8-10].

Medication nonadherence is a particularly problematic issue affecting older adults and results in suboptimal control of chronic conditions leading to poor health outcomes, hospitalizations, and significant health care costs [11,12]. For example, among patients aged 75 years and older, nonadherence to bisphosphonates resulted in an odds ratio of 1.49 for osteoporotic fractures and 13.4% higher medical costs than adherent patients [13]. Among older adults with epilepsy, the odds ratio of utilization of inpatient services was 0.66 among patients adherent to antiepileptic drugs compared with nonadherent patients and 13.2% less cost on the total direct health care costs [14]. Similarly, a population-level evaluation of the economic impact of nonadherence to medications used for the treatment of diabetes, heart failure, hyperlipidemia, and hypertension among Medicare beneficiaries in the United States demonstrated that improvement in adherence would result in an annual reduction of 117, 594 emergency department visits and over 7 million inpatient hospital days among patients with hypertension [15]. Likewise, reductions in these health care costs would also be realized in patients with hyperlipidemia, diabetes, and heart failure [15].

Adherence is defined as “the degree to which the person’s behavior corresponds with the agreed recommendations of a healthcare provider” [16]. Nonadherence to medications can arise from an intentional, active, and conscious decision to deviate, or result from passive, unintentional divergence from the prescribed dosing regimen [17,18]. Common causes of intentional nonadherence among older adults include beliefs related to the patient perception of illness and necessity for medications, experience of adverse events from medications, patient-prescriber relationship, and complexities related to dosing regimens of multiple medications [17]. Among older adults, nonadherence to medications can arise from several different factors, including patient-related, medication-related, health care provider–related, health care system–related, and socioeconomic-related factors [19]. Among the medication-related factors that can lead to inadvertent or unintentional nonadherence are dosing regimens, formulations, and packaging of medications [19]. Additionally, patient-related factors such as physical functioning (eg, vision limitations, impaired hearing, and poor dexterity) and cognitive impairment impact a patient’s ability to accurately administer medications, that is, the ability to take the right dose of the right medication at the right time [19-27]. These limitations are well recognized in both the clinical and research environments. As a result, several tools that measure the functional capacity of an older adult to manage medications have been developed to assist clinicians and patients in identifying different limitations in physical and cognitive capacities [28,29]. Several different strategies have been designed to address this need, including telephone reminders, memory aids and cues, pill boxes and dosettes, and compliance or blister-packaging, among others. However, systematic reviews indicate that the effectiveness of many of these strategies has produced mixed results in terms of the impact on adherence [30-32]. A systematic search for electronic products marketed to assist with medication management conducted in 2016 revealed that more than 80 such products were available for purchase to Canadians to address medication management needs [33]. These electronic medication adherence products may have audio and/or visual alarms, locking features, report-generating abilities, and real-time adherence monitoring, among others. These electronic medication adherence products are physical products that a patient can purchase and are not mobile or web-based apps. They may have a mobile or web-based app accessory to the product, but this is not the sole product. Very few of these commercially available electronic medication adherence products have been tested for usability and/or workload.

The usability of a product refers to the “facility with which users can use a technological artefact to achieve a particular goal” [34]. Before examining the effectiveness of these products on adherence and clinical outcomes, usability should be established. Products that are not user-friendly have the potential to worsen adherence in older adults. If a product is not practical, is socially unacceptable, has limited learnability, is inefficient and unpleasant to use, or results in increasing error rates, it may negatively affect adherence [35]. Usability is especially imperative to be established among older adults, given the higher prevalence of cognitive and physical impairments. A conceptual framework developed to address the usability of health technology among older adults highlights 4 different categories, including cognitive barriers, physical impairments, motivational issues, and perception barriers that impact usability and must be addressed [36]. However, usability of the majority of electronic medication adherence products has not been established. Of the products that have been tested, usability varies based on its features and limitations encountered in the participant population. In one study, 96 frail older adults used an electronic automated medication dispensing device for 1 year [37]. In this study, 94% (90/96) of the participants found the device to be easy to use, 95% (91/96) reported that it was reliable, and 95% (91/96) reported that the dispenser gave them peace of mind. However, only 84% (81/96) would consider using the dispenser in the future, and 5 participants indicated that they would require continued assistance from a nurse with pill setup. It is important to note that during this study, participants were not required to fill the dispenser with medications. In another study, 90% (18/20) of the 20 participants aged 55-75 years also indicated satisfaction with a personalized medication support system developed for Android platforms using quick response codes. However, users requested further improvements in printed characters, font sizes, and compatibility, and they noted financial constraints and the requirement of smartphones as limitations [38]. Similar limitations were identified in another study designed to investigate the usability of EMMA (R) among older adults [39]. Major challenges with usability were identified, including a narrow medication loading slot; difficulty in reading the font; and difficulty in identifying, retrieving, and opening delivered medications. In a separate study, usability of medication adherence technology was found to be limited in persons with cognitive impairment [40]. In this study, Mini-Mental Status Examination (MMSE) scores were significantly correlated with the percentage of task success; noncognitively impaired individuals completed 69% of the tasks required to use an integrated medication unit, whereas cognitively impaired individuals were only successful at completing 34.7% of the tasks [40]. Similar issues with navigation, poor visibility, and lack of transparency have been identified in medication management apps available for use in smartphones [41,42]. Usability scores of some of these apps ranged from 42 to 57 on the system usability score (SUS), a validated measure of usability where products are scored from 0 to 100, with higher scores indicating better usability [41,42].

Objectives

Given that these studies demonstrated significant variability and limitations in the usability of health technology addressing medication management capacity and adherence and that the usability of many commercially available electronic medication adherence products has not been tested, this study aimed to examine the usability and workload of a range of electronic medication adherence products among older adults, caregivers, and health care providers.

Methods

Study Design

This study was designed as a prospective, mixed methods study to investigate the usability and workload of 21 electronic medication adherence products. In an effort to measure usability in a multidomain approach, we used both quantitative measures, including validated tools, cognitive walkthroughs, observations, and qualitative one-on-one interviews [35,43]. The paper describes the quantitative findings of this study. We did not develop any of the products tested in our study.

Sample and Sample Size

We used purposive sampling techniques to recruit three types of participants for this study: older adults, caregivers of older adults, and health care providers. To be eligible, older adults had to be aged 65 years and older and taking one or more prescription or nonprescription medications regularly, whereas caregivers had to be assisting older adults with medication administration. Older adults and caregivers were recruited through advertising, professional networks, and from a participant pool of older adults who had previously indicated an interest in participating in research. All health care providers were eligible for the study and were recruited primarily through professional networks of researchers.

As 80% of the usability problems can be identified with 5 test-users [34], we aimed to test each product with at least five participants. Therefore, we targeted a recruitment sample of 25, composed of 15 older adults or caregivers and 10 health care providers.

Products Included

In total, 23 electronic medication adherence products were purchased for the purposes of this study; however, 1 product was nonfunctional and, as such, was not tested. One product was found to be nonelectronic; however, it was still tested, although the results are not included in this report. Products were chosen with the objective of representing different features such as automated dispensing, number and type of compartments, audio and vibration reminder alarms, cloud connectivity, and medication dispensing tracking. Products that cost over Can $1000 (US $710.84) to purchase were not selected based on funding constraints. The names and details of the electronic medication adherence products tested are described in Table 1.

Table 1. Electronic medication adherence products tested.

Name of product	Abbreviated name	Manufacturer	Price (US $)
GMS[a] Med-e-lert automatic pill dispenser[b]	Med-e-lert	Group Medical Supply, LLC[c]	70-109
LiveFine automatic pill dispenser and reminder [b]	LiveFine	LiveFine	70-109
MedReady 1700 automated medication dispenser[b]	MedReady	MedReady Inc	≥109
MedSmart med-reminder and dispensing system	MedSmart	e-pill[d]	≥109
e-pill MedTime Station automatic pill dispenser with tipper	MedTime Station	e-pill	≥109
TimerCap travel size	TimerCap T	TimerCap, LLC	<30
TimerCap universal size	TimerCap U	TimerCap, LLC	<30
Jones medication adherence system 14-unit card	Jones	Jones Packaging Inc	N/A
Reizen vibrating pill box	Reizen	Reizen	<30
VitaCarry advanced pill case[b]	VitaCarry	VitaCarry	30-69
Nishiki round pill box with alarm[b]	Nishiki	Nishiki	<30
MedGlider system 1 with talking reminder[b]	Medport MedGlider	Medport	30-69
Patterson medical tabtime super 8[e]	TabTime	Tabtime LTD[f]	30-69
100-Hour pill reminder[e]	100-Hour	Aidapt	<30
Med-Q smart pill box	Med-Q	Med-Q	70-109
e-pill MedGlider home medication management system	e-pill MedGlider	e-pill	30 - 69
MedCentre system[b]	MedCentre	MedCenter Systems, LLC	30-69
eNNOVEA Weekly Planner With Advanced Auto Reminder	eNNOVEA	eNNOVEA Medical, LLC	70-109
e-pill Multi-alarm pocket XL	Multi-Alarm	e-pill	30-69
6 Grid pill storage case with alarm	Pill Storage Case	NR	<30
Itzbeen pocket doctor[b]	Pocket Doctor	Itzbeen	<30

[a] GMS: Group Medical Supply.

[b] Purchased from Amazon Canada or the United States.

[c] LLC: Limited Liability Company.

[d] e-pill: electronic pill.

[e] Purchased from Ebay Canada.

[f] LTD: Limited.

Outcome Measures

Usability: System Usability Scale

SUS is a popular end-of-test subjective assessment of the usability of a product [44,45]. It is a validated, quick, and easily administered tool that has been used to test products across a wide range of industries. It consists of 10 statements (5 positive and 5 negative), which are scored immediately after testing a product on a 5-point Likert scale. Scores range from 0 to 100, with higher scores indicating that a product is more usable.

Workload: NASA-Task Load Index

The NASA-task load index (NASA-TLX) measures the workload required to complete a task [46]. It consists of 6 subscales: mental, physical, demand, frustration, effort, and performance. Participants were asked to rate each of the above-mentioned variables on a 20-point scale that measures from high to low (scored from 0 to 100) [46]. The tool has been applied in various fields, such as measuring workload in persons working in critical care, performing surgery, commercial aviation, and daily activities (such as operating a medical device at home) [47]. There are 2 ways to calculate the score. For the purpose of this study, we calculated the average of the ratings of the 6 items for each participant [47]. Lower scores indicate less workload.

Task Completion Times, Task Completion Rates, and Error Rates: Cognitive Walkthrough Checklist, Thinking Aloud

A cognitive walkthrough checklist was used to portray the cognitive tasks an individual performs while completing a series of complex tasks [48]. Individual cognitive walkthrough checklists were developed for each tested electronic medication adherence product. Cognitive walkthroughs were designed as outlined by Kushrinuk and Patel [48] and included (1) identifying the end users of the medication adherence products (older adults, caregivers, and health care providers), (2) defining the tasks for the walkthrough based on instructions provided by the manufacturer and supplemental information available on the web, and (3) walking through the actions and critiquing the vital information and steps with several pilot runs by research team members. Any critical steps, defined as those that were necessary for the appropriate use of the product, were identified and noted. The cognitive walkthrough checklist for each product listed each task, the steps required to complete the task, space to note errors performed in completing the tasks as well as time to complete, and if the completion of all the tasks was unassisted or partially assisted. The tasks varied for each product, but generally included steps such as opening and filling a tray with medications by using a mock medication regimen with placebo tablets and capsules, setting an alarm, locking a device, and removing the placebo tablets or capsules. As participants tested each product, they were invited to *think aloud*, that is, verbalize their thoughts as they completed the tasks required to use a medication adherence product with the mock medication regimen [48]. The *think aloud* sessions were audio recorded using a digital recorder (Sony IC Recorder ICD-PX470), while also being observed by a research team member who recorded their observations on the cognitive walkthrough checklist. Having the patient verbalize their thoughts helped researchers gain insight into what the participant was thinking and why certain errors were occurring. Participants were timed from the point at which they started testing the adherence product up to the point at which they completed all the tasks required to use the product or the point at which the participant refused to continue because of frustration, confusion, or fatigue with completing the remaining tasks. Timing was measured using an Apple iPhone X (software version iOS 13.31, 2017) clock app with a built-in timer.

Mock Medication Regimen

A mock polypharmacy medication regimen was developed for the purpose of this study. Placebo tablets (national product number 00501190, Odan Laboratory, lot #188628A, 100 mg lactose tablets), candy (Tic Tac), and placebo capsules (manufactured in-house) were used to represent the following medications (with administration instructions as indicated here): warfarin 2 mg once daily on Monday, Wednesday, and Friday and 3 mg once daily on Tuesday, Thursday, Saturday, and Sunday; pantoprazole 20 mg twice daily; phenytoin 100 mg, 1 capsule in the morning and 2 capsules in the evening; and propranolol 20 mg, half a tablet once daily for 2 days, then 1 tablet daily. The dosing regimen was designed to reflect commonly prescribed regimens. Participants were required to use the mock medications and instructions provided on the prescription labels to fill the medication adherence products they were testing in an effort to reflect activities conducted in real-life scenarios. For each product participants tested, they were required to use the same mock regimen.

Testing Procedures

Each participant tested 5 electronic medication adherence products. No training was provided to participants as this option would not be available in the community setting, but participants were provided with manufacturer instructions. If manufacturer instructions were not enclosed with the product, researchers accessed web-based instructions and provided them to the participant. Each participant was required to perform a series of tasks while *thinking aloud*, covering all aspects of using the product. At the end of each product testing, participants were asked to complete both the SUS and NASA-TLX. Finally, each participant was invited to an optional one-to-one semistructured interview examining participant perceptions about the features of the products, if they would consider using the product for managing their medications or if they would recommend the product to friends or family members. The results of the qualitative analysis for the one-to-one interviews are not presented in this analysis.

Data Collection and Statistical Analysis

The SUS and NASA-TLX were provided to participants to complete in a paper format at the end of each product testing. The cognitive walkthrough checklist was completed by a research team member for each product. The audio recordings were then used to ensure the accuracy of the observations completed by a second team member. Data were entered in Microsoft Excel spreadsheets (Microsoft Excel version 16.16.14, 2016), Microsoft Word (Microsoft Word version 16.16.13, 2016), and Microsoft Access (Microsoft Access Version 16.0.4738.100, 2016).

SUS and NASA-TLX scores for the different electronic medication adherence products were analyzed using RStudio version 3.5.1 (2018-07-02). A repeated-measures analysis of variance (ANOVA) was used to compare both SUS and NASA-TLX scores of all the products, followed by Tukey posthoc analysis. The Pearson correlation was used to determine if there was a statistically significant relationship between SUS scores, that is, usability and NASA-TLX scores, that is, workload.

Error rate was calculated by dividing the number of total errors made (per person per product) by the total number of steps required to use the product (per product). Mean error rates are reported as percentages. Unassisted task completion rates were calculated by dividing the total number of tasks completed without assistance by each participant for each product by the total number of tasks required to use the particular product. The means of this measure are provided as percentages. Completion rates were calculated as the total number of tasks completed with or without assistance divided by the total number of tasks. Completion times reported are the mean time required to complete the setup and use of each product per participant. Error rates, completion times, and completion rates for each electronic medication adherence product tested are summarized.

Ethical Review and Location

This study was reviewed by and received approval from the University of Waterloo Office of Research Ethics. All participants were informed of the study and provided consent before enrolling. This study was conducted at the University of Waterloo School of Pharmacy in Kitchener, Ontario, Canada.

Results

Participant Demographics

A sample of 39 individuals were recruited to test the electronic medication adherence products in this study, of which 23 were older adults, 5 were caregivers, and 11 were health care professionals. The majority of the 23 older adults were taking more than 5 medications, including prescriptions, vitamin supplements, over the counter drugs, and/or natural health products concurrently (see Table 2).

Almost 70% (16/23) of older adults also reported using a medication aid to assist with medication management, with the most commonly utilized being pill boxes, followed by alarms on a phone/watch and blister pack. All 5 caregivers assisted older adults with their medication management. Of the 11 health care providers, all of them worked with older adults in some capacity, with more than 80% (9/11) assisting with medication taking activities, but all recommending some type of medication aid to older adults (see Table 3).

Table 2. Demographic characteristics of older adults and caregivers.

Variable	Older adults (n=23)	Caregivers(n=5)
Gender, n (%)		
Male	11 (48)	4 (80)[a]
Female	12 (52)	1 (20)[a]
Age (years)		
Mean (SD)	75 (6.7)	73.2 (4.49)[a]
Mode	82	N/A[b]
Median	75	71[a]
Range	65-87	69-79[a]
Number of medications taken per participant		
Mean (SD)	7.5 (3.3)	9.8 (5.3)[a]
Mode	8	N/A
Median	8	8.5[a]
Range	1-13	5-17[a]
Total number of drugs	172	39[a]
Number of participants taking more than 5 medications per day, n (%)		
Prescription medications, OTC[c], vitamin supplements, and herbal	19 (83)	—[d]
Prescription medications	10 (44)	—
OTC, vitamin supplements, and herbal	6 (26)	—
Medication schedule, n (%)		
Once daily	4 (17)	0 (0)[a]
More than once daily	19 (83)	5 (100)[a]
Medication aids use, n (%)		
Yes		
Total	16 (70)	5 (100)[a]
Blister pack	3 (13)	2 (40)[a]
Pill box	12 (52)[e]	4 (80)[a]
Alarm (phone and/or watch)	4 (17)	0 (20)[a]
App	1 (4)	0 (20)[a]
Calendar and/or message board and/or list	2 (9)	1(20)[a]
No	7 (30)	0 (0)[a]
Medication aids used in combination, n (%)		
Yes		
Total	5 (22)	2 (40)[a]
Pill box and alarm	2 (9)	0 (0)[a]
Pill box and message board and/or list	2 (9)	1 (20)[a]
Pill box and app	1 (4)	0 (0)[a]
Pill box and blister pack	0 (0)	1 (20)[a]

[a] Reported by caregiver for the patient.

[b] N/A: not applicable.

[c] OTC: Over the counter.

[d] Not reported.

[e] One participant reported not using any medication aid but stated that they used a pill case only for traveling purposes.

Table 3. Demographic characteristics of health care providers.

Variable	Health care providers (n=11)
Gender, n (%)	
Male	2 (18)
Female	9 (82)
Occupation, n (%)	
Pharmacist	8 (72)
Pharmacy student	1 (9)
Occupational therapist	2 (18)
Years of practice	
Mean (SD)	8.8 (10.5)
Mode	1
Median	5
Range	0[a]-37
Older adults worked with/dispensed prescriptions for, n (%)	
<10	1 (9)
10-20	4 (36)
20-30	1 (9)
>30	5 (46)
Assist older adults with medication taking, n (%)	
Yes	9 (82)
No	2 (18)
Medication aids recommendation, n (%)	
Yes	
Total	11 (100)
Blister pack	10 (91)
Pill box/dosette	6 (55)
Easy snap cap	1 (9)
Alarm	3 (27)
Phone app	1 (9)
No	0 (0)

[a] One health care provider was a pharmacy student and thus had 0 years of practice as a registered pharmacist.

Usability

The overall mean SUS score for all 21 products tested was 52.28 (SD 28.52; range 0-100). The mean SUS score per product ranged from a low of 28.63 (SD 21.240) for the MedReady 1700 automated medication dispenser to a high of 78.67 (SD 15.572) for the TimerCap travel size (Table 4). The mean SUS scores were significantly different between MedReady 1700 automated medication dispenser and the TimerCap travel size (mean 28.63, SD 21.24 vs mean 78.67, SD 15.57; *P*=.03) and MedReady 1700 automated medication dispenser and the eNNOVEA Weekly Planner with Advanced Auto Reminder (mean 28.63, SD 21.24 vs mean 78.13, SD 14.13; *P*=.049).

Table 4. Mean system usability scale and NASA-task load index scores

Product name	System usability scale		NASA-task load index	
	Score, mean (SD)	Range	Score, mean (SD)	Range
Med-e-lert	40.75 (31.78)	7–85	67.08 (25.60)	33.33–97.50
LiveFine	46.50 (27.38)	10–85	65.21 (21.32)	32.50–99.17
MedReady	28.63 (21.24)	0–65	72.92 (18.69)	43.33–92.50
MedSmart	41.44 (33.97)	5–87	59.23 (25.82)	13.33–93.33
MedTime Station	35.13 (17.84)	20–72	71.77 (21.98)	2.67–99.17
TimerCap T	78.67 (15.57)	50–100	29.35 (20.35)	4.17–55.83
TimerCap U	56.11 (30.92)	0–92	40.56 (27.72)	12.5–99.17
Jones	50.64 (24.63)	7–92	48.63 (18.11)	23.33–85.83
Reizen	50.67 (24.06)	20–92	50.28 (27.46)	4.17–76.67
VitaCarry	52.78 (31.69)	0–87	46.11 (23.45)	13.33–83.33
Nishiki	46.89 (26.50)	0–90	56.48 (26.12)	11.67–97.5
Medport MedGlider	57.20 (30.85)	0–95	45.75 (24.05)	18.33–95.00
TabTime	57.89 (31.36)	0–97	35.65 (20.30)	4.17–68.33
100-Hour	55.78 (25.28)	2–85	49.35 (24.57)	26.67–95.83
Med-Q	64.60 (27.85)	12–90	42.17 (28.70)	14.17–89.17
e-pill MedGlider	71.67 (30.84)	5–100	28.43 (20.80)	4.17–69.17
MedCentre	56.63 (22.10)	37–92	52.19 (18.52)	29.17–79.17
eNNOVEA	78.13 (14.13)	55–97	28.65 (14.97)	10.00–57.5
Multi-Alarm	51.56 (24.59)	15–95	45.65 (25.79)	4.17–78.33
Pill Storage Case	34.75 (22.60)	0–62	65.10 (14.17)	41.67–80.83
Pocket Doctor	35.00 (33.19)	0–92	68.33 (27.37)	20.00–93.33

Workload

For the 21 products tested, the overall mean NASA-TLX score was 50.43 (SD 25.49; range 4.17-99.17). The mean NASA-TLX score per product ranged from 28.65 (SD 14.97) for the eNNOVEA Weekly Planner with Advanced Auto Reminder to 72.92 (SD 18.69) for the MedReady 1700 automated medication dispenser (Table 4). Similar to SUS scores, ANOVA analysis revealed significant differences in the NASA-TLX raw scores. The NASA-TLX raw scores for the MedReady 1700 automated medication dispenser (mean 72.92, SD 18.69) were significantly higher than those for the TimerCap travel size (mean 29.35, SD 20.35; *P*=.03), e-pill MedGlider home medication management system (mean 28.43, SD 20.80; *P*=.02), and eNNOVEA (mean 28.65, SD 14.97; *P*=.03). The e-pill MedTime Station automatic pill dispenser with tipper (mean 71.77, SD 21.98) demonstrated significantly higher workload than TimerCap (*P*=.04), MedGlider (*P*=.03), and eNNOVEA (*P*=.04).

Correlation Between System Usability Scale Scores and NASA-Task Load Index Scores

The SUS scores were highly correlated with NASA-TLX scores (Pearson $r=-0.877$; 95% CI −0.907 to 0.839; $df=184$; $P<.001$), such that an increase in usability of the product was inversely correlated with decreasing workload involved in using the product.

Mean Time to Task Completion, Task Completion, and Error Rates

The number of tasks and the number of steps within each task to use a product varied depending on the characteristics and features of the products. Products required an average of 12.7 steps (SD 4.1; range 5-20) for use and included steps such as inserting the key into the lock, rotating the key clockwise, and pinching the tab to open the product. These steps were grouped into tasks such as opening the product, loading the product with placebo pills or Tic Tac candy as per the standardized mock medication schedule provided, closing the product, setting an alarm, and removing the medications from the product after the alarm sounds (Table 5). On average, with assistance from a research team member, participants were able to complete 97% of the tasks required to use a product (SD 8.97%; range 16.7%-100%). However, without the assistance of a researcher, participants were only able to complete 100% of all steps required to use a product in 103 of 186 product tests (103/168, 55.3%; range: 0%-100%).

Table 5. Mean error rates, completion time, unassisted task completion, and completion rates of electronic medication adherence products tested.

Abbreviated product name	Total number of steps	Description of tasks					Mean time to task completion (min:sec)	Mean error rate (%)	Mean unassisted task completion (%)	Mean completion rate (%)
		Open and fill medication tray	Set alarm	Lock device	Remove medication from device	Other				
Med-e-lert	18	X[a]	X	X	X	N/A[b]	17:24	22	81	98
LiveFine	18	X	X	X	X	N/A	18:22	17	70	90
MedReady	16	X	X	X	X	N/A	26:15	21	86	98
MedSmart	20	X	X	X	X	N/A	25:20	19	67	98
MedTime Station	17	X	X	N/A	X	X	31:13	36	66	98
TimerCap T	5	X	X	N/A	N/A	X	5:36	2	100	100
TimerCap U	5	X	X	N/A	N/A	X	5:19	7	100	100
Jones	15[c]	N/A	N/A	N/A	X	X	15:17	35	80	92
Reizen	10	X	X	N/A	X	N/A	15:29	15	78	95
VitaCarry	10	X	X	N/A	X	N/A	15:11	20	88	99
Nishiki	10	X	X	N/A	X	N/A	15:10	26	82	98
Medport MedGlider	11	X	X	N/A	X	N/A	16:16	10	95	100
TabTime	12	X	X	N/A	X	N/A	12:20	16	81	99
100-Hour	10	X	X	N/A	X	N/A	9:16	3	96	100
Med-Q	12	X	X	N/A	X	N/A	12:16	27	85	100
e-pill MedGlider	14	X	X	N/A	X	N/A	10:13	17	90	100
MedCentre	15	X	X	N/A	X	N/A	16:23	10	72	94
eNNOVEA	14	X	X	N/A	X	X	15:31	10	88	98
Multi-Alarm	12	X	X	N/A	X	N/A	11:22	25	78	97
Pill Storage Case	12	X	X	N/A	X	N/A	15:32	24	79	95
Pocket Doctor	17	N/A	X	N/A	X	X	15:11	24	68	85

[a] Task needed to be completed.

[b] N/A: not applicable.

[c] A total of 3 tasks (of the total 15) were only applicable for health care providers to complete. Thus, the total number of steps is 12 for an older adult or caregiver participant, and it is 15 for a health care provider participant.

Discussion

Principal Findings

To our knowledge, our study is the first and the largest study to examine and compare the usability and workload of electronic medication products in a population of older adults, caregivers, and health care providers. We tested a range of products, varying between products with few features to complex devices with multiple features, using a standardized mock medication regimen that mimicked polypharmacy and medication regimen complexity that is prevalent in an older adult population [5]. The electronic medication adherence products we purchased were discovered in a systematic search we conducted in a previous study and reflected a variety of characteristics and available features [33]. The results of our study demonstrate significant variability in both the usability and workload required to use electronic medication adherence products. Additionally, the task completion rates varied between electronic medication adherence products as well as the number of task errors.

We examined the usability of electronic medication adherence products by older adults, caregivers, and health care providers. The results of this study are of importance to both older adults and health care providers, especially when determining which electronic medication adherence product can be used to address medication management at home. Higher usability scores indicate that an older adult, caregiver, or health care provider is more likely to be able to use a product appropriately, potentially decrease medication error rates, and improve adherence. Products with lower usability scores may introduce dosing errors, reduce adherence, and, at worst, introduce medication errors. However, although the results of this study demonstrate that some products score well for usability, making them more user-friendly, and others score poorly demonstrating poor usability, there is no benchmark for SUS scores for electronic medication adherence products or for any adherence products for that matter. Therefore, an exact interpretation of this score is difficult. In comparison, the mean SUS scores for other everyday products are as follows: Excel, 56.5; iPhone, 78.5; and microwaves, 86.9 [45,46,48,49]. As a cutoff score above which a product could be deemed highly usable is not available, Bangor et al [45] described an emergence of scoring systems using adjectives such as *acceptable* for products that score above 70 on SUS, *marginal* for scores of 50-69, and *not acceptable* for scores of less than 50. It is easier to relate the product's usability with the use of these descriptive terms in clinical practice, especially when determining the potential usability of a product. In terms of the products we tested, we would then consider the TimerCap travel size, e-pill MedGlider home medication system, and the eNNOVEA Weekly Planner with Advanced Auto Reminder as products which have *acceptable* usability, whereas the TimerCap universal size, Jones medication adherence system, Reizen vibrating pill box, VitaCarry advanced pill case, MedGlider system 1 with talking reminder, Patterson medical TabTimer super 8, 100-Hour pill reminder, Med-Q smart pill box, MedCentre system, and e-pill multi-alarm pocket XL as *marginally acceptable*, and the remaining products as *unacceptable* for user-friendliness. This type of classification would allow clinicians to evaluate the potential problems with usability their older adult population would encounter. It also allows older adults and caregivers to evaluate products before making a financial investment.

Similar to SUS, the NASA-TLX has been applied in various fields. A meta-analysis of NASA-TLX scores in over 200 publications reported an overall range of scores from 6.21 to 88.5 for various activities (eg, air traffic control, cognitive activities, daily activities, and driving care) [47]. As with the SUS scores, no benchmark for the NASA-TLX regarding medication adherence products, electronic or not, is available. This limits the ability to generalize the scores to a level of workload, that is, to determine the cutoffs for scores that could be used to designate specific products as excellent, moderate, or poor, as has been recommended with SUS scores. Furthermore, adjectives such as those that exist with SUS are not available to allow us to easily apply the scoring in clinical practice. However, the NASA-TLX scores can be compared between the different products to gauge if the workload required to use a product in a population such as ours would be high or not, relative to other products tested. Additionally, by plotting the scores of SUS and NASA-TLX for each product, it became apparent that they were inversely related, that is, products that scored high for usability also scored low for workload. We were also able to demonstrate that the 2 variables are highly correlated (Pearson's $r=-0.877$; 95% CI −0.907 to 0.839; $df=184$; $P<.001$). Given this finding, it is reasonable to expect that a product that scores well on usability does not place an undue workload cost on the human user.

Strengths of the Study

Older adults rarely report the problems they have with medication management, and often, they develop their own strategies to address these problems, including the use of commercially available electronic medication adherence products. Our study is the first to compare a number of these products, and the results from our study indicate that usability, workload, time taken to complete tasks, and task completion rates vary between different products. The results of our study provide some evidence for factors to consider when choosing electronic medication adherence products in older adults at risk of medication nonadherence and can also help clinicians and caregivers to guide the choice of an electronic medication adherence product. Although much work needs to be done to investigate the impact of product choice on adherence and clinical outcomes, our study provides evidence to guide the first step in the process, that is, the choice of the product to use.

Limitations of the Study

This study has some limitations. The participants tested these products in an unfamiliar environment (the University of Waterloo School of Pharmacy) while being observed by research team members, which may have produced or increased anxiety in some of the participants and impacted their ability to use the products efficiently and effectively. Furthermore, participants were asked to test a product they had not been previously exposed to, and this may have impacted the scoring; gaining familiarity with products usually improves the ease of use. Participants were also asked to test 5 products, and frustration with one or more products may have filtered over to the other products they were asked to test thereafter.

Conclusions

As few studies have investigated the usability of electronic medication adherence products, benchmarks do not yet exist to compare the usability and workload of these products. Our study showed that some electronic medication adherence products may be easier to use than others. Furthermore, information such as usability and workload will provide older adults and clinicians with factors to consider when recommending or purchasing devices to address medication management among older adults.

Acknowledgments

This study was supported by the Spark Program at the Centre for Aging and Brain Health Innovation. The authors thank all the participants in the study for their time and input. The authors are grateful to Catherine Lee for assisting with copy editing and formatting the manuscript.

Conflicts of Interest

None declared.

References

1. United Nations. World Population Ageing 2017 Highlights. New York City, USA: United Nations; 2017.
2. Suzman R, Beard J. World Health Organization. 2011. Global Health and Aging URL: https://www.who.int/ageing/publications/global_health.pdf [accessed 2020-04-15]
3. Salive ME. Multimorbidity in older adults. Epidemiol Rev 2013;35:75-83. [doi: 10.1093/epirev/mxs009] [Medline: 23372025]
4. Vasilopoulos T, Kotwal A, Huisingh-Scheetz MJ, Waite LJ, McClintock MK, Dale W. Comorbidity and chronic conditions in the national social life, health and aging project (NSHAP), wave 2. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69(Suppl 2):S154-S165 [FREE Full text] [doi: 10.1093/geronb/gbu025] [Medline: 25360017]
5. Canadian Institute for Health Information. Drug Use Among Seniors in Canada, 2016. Ottawa, ON: CIHI; 2018.
6. Charlesworth CJ, Smit E, Lee DS, Alramadhan F, Odden MC. Polypharmacy among adults aged 65 years and older in the United States: 1988-2010. J Gerontol A Biol Sci Med Sci 2015 Aug;70(8):989-995 [FREE Full text] [doi: 10.1093/gerona/glv013] [Medline: 25733718]
7. Payne RA. The epidemiology of polypharmacy. Clin Med (Lond) 2016 Oct;16(5):465-469 [FREE Full text] [doi: 10.7861/clinmedicine.16-5-465] [Medline: 27697812]
8. Maher RL, Hanlon J, Hajjar ER. Clinical consequences of polypharmacy in elderly. Expert Opin Drug Saf 2014 Jan;13(1):57-65 [FREE Full text] [doi: 10.1517/14740338.2013.827660] [Medline: 24073682]
9. Leelakanok N, Holcombe A, Lund B, Gu X, Schweizer M. Association between polypharmacy and death: a systematic review and meta-analysis. J Am Pharm Assoc (2003) 2017;57(6):729-38.e10. [doi: 10.1016/j.japh.2017.06.002] [Medline: 28784299]
10. Salvi F, Marchetti A, D'Angelo F, Boemi M, Lattanzio F, Cherubini A. Adverse drug events as a cause of hospitalization in older adults. Drug Saf 2012 Jan;35(Suppl 1):29-45. [doi: 10.1007/BF03319101] [Medline: 23446784]
11. Chiatti C, Bustacchini S, Furneri G, Mantovani L, Cristiani M, Misuraca C, et al. The economic burden of inappropriate drug prescribing, lack of adherence and compliance, adverse drug events in older people: a systematic review. Drug Saf 2012 Jan;35(Suppl 1):73-87. [doi: 10.1007/BF03319105] [Medline: 23446788]
12. Butler R, Davis T, Johnson W, Gardner H. Effects of nonadherence with prescription drugs among older adults. Am J Manag Care 2011 Feb;17(2):153-160 [FREE Full text] [Medline: 21473664]
13. Moser SS, Yu J, Goldshtein I, Ish-Shalom S, Rouach V, Shalev V, et al. Cost and consequences of nonadherence with oral bisphosphonate therapy: findings from a real-world data analysis. Ann Pharmacother 2016 Apr;50(4):262-269. [doi: 10.1177/1060028015626935] [Medline: 26783359]
14. Ip Q, Malone DC, Chong J, Harris RB, Labiner DM. Economic impact of epilepsy and the cost of nonadherence to antiepileptic drugs in older medicare beneficiaries. Epilepsy Behav 2018 Mar;80:208-214. [doi: 10.1016/j.yebeh.2018.01.009] [Medline: 29414554]
15. Lloyd JT, Maresh S, Powers CA, Shrank WH, Alley DE. How much does medication nonadherence cost the medicare fee-for-service program? Med Care 2019 Mar;57(3):218-224. [doi: 10.1097/MLR.0000000000001067] [Medline: 30676355]
16. World Health Organization. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2013.
17. Wroe AL. Intentional and unintentional nonadherence: a study of decision making. J Behav Med 2002 Aug;25(4):355-372. [doi: 10.1023/a:1015866415552] [Medline: 12136497]
18. Mukhtar O, Weinman J, Jackson SH. Intentional non-adherence to medications by older adults. Drugs Aging 2014 Mar;31(3):149-157. [doi: 10.1007/s40266-014-0153-9] [Medline: 24566876]
19. Yap AF, Thirumoorthy T, Kwan YH. Systematic review of the barriers affecting medication adherence in older adults. Geriatr Gerontol Int 2016 Oct;16(10):1093-1101. [doi: 10.1111/ggi.12616] [Medline: 26482548]
20. Mullen R, Curtis L, O'Conor R, Serper M, McCarthy D, Bailey S, et al. Visual acuity, literacy, and unintentional misuse of nonprescription medications. Am J Health Syst Pharm 2018 May 1;75(9):e213-e220. [doi: 10.2146/ajhp170303] [Medline: 29691264]
21. Philbert D, Notenboom K, Bouvy ML, van Geffen EC. Problems experienced by older people when opening medicine packaging. Int J Pharm Pract 2014 Jun;22(3):200-204. [doi: 10.1111/ijpp.12070] [Medline: 24025086]
22. Sino CG, Sietzema M, Egberts TC, Schuurmans MJ. Medication management capacity in relation to cognition and self-management skills in older people on polypharmacy. J Nutr Health Aging 2014 Jan;18(1):44-49. [doi: 10.1007/s12603-013-0359-2] [Medline: 24402388]
23. Notenboom K, Beers E, van Riet-Nales DA, Egberts TC, Leufkens HG, Jansen PA, et al. Practical problems with medication use that older people experience: a qualitative study. J Am Geriatr Soc 2014 Dec;62(12):2339-2344 [FREE Full text] [doi: 10.1111/jgs.13126] [Medline: 25516030]
24. Atkin PA, Finnegan TP, Ogle SJ, Shenfield GM. Functional ability of patients to manage medication packaging: a survey of geriatric inpatients. Age Ageing 1994 Mar;23(2):113-116. [doi: 10.1093/ageing/23.2.113] [Medline: 8023717]

25. Spiers M, Kutzik D, Lamar M. Variation in medication understanding among the elderly. Am J Health Syst Pharm 2004 Feb 15;61(4):373-380. [doi: 10.1093/ajhp/61.4.373] [Medline: 15011765]
26. Cárdenas-Valladolid J, Martín-Madrazo C, Salinero-Fort MA, de-Santa Pau EC, Abánades-Herranz JC, de Burgos-Lunar C, PATER Group. Prevalence of adherence to treatment in homebound elderly people in primary health care: a descriptive, cross-sectional, multicentre study. Drugs Aging 2010 Aug 1;27(8):641-651. [doi: 10.2165/11537320-000000000-00000] [Medline: 20658792]
27. Maddigan SL, Farris KB, Keating N, Wiens CA, Johnson JA. Predictors of older adults' capacity for medication management in a self-medication program: a retrospective chart review. J Aging Health 2003 May;15(2):332-352. [doi: 10.1177/0898264303251893] [Medline: 12795276]
28. Farris KB, Phillips BB. Instruments assessing capacity to manage medications. Ann Pharmacother 2008 Jul;42(7):1026-1036. [doi: 10.1345/aph.1G502] [Medline: 18594054]
29. Elliott RA, Marriott JL. Standardised assessment of patients' capacity to manage medications: a systematic review of published instruments. BMC Geriatr 2009 Jul 13;9:27 [FREE Full text] [doi: 10.1186/1471-2318-9-27] [Medline: 19594913]
30. Gellad WF, Grenard JL, Marcum ZA. A systematic review of barriers to medication adherence in the elderly: looking beyond cost and regimen complexity. Am J Geriatr Pharmacother 2011 Feb;9(1):11-23 [FREE Full text] [doi: 10.1016/j.amjopharm.2011.02.004] [Medline: 21459305]
31. Cross A, Elliott R, George J. Interventions for improving medication-taking ability and adherence in older adults prescribed multiple medications. Cochrane Database Syst Rev 2016;2016(10):-. [doi: 10.1002/14651858.CD012419/full]
32. Costa E, Giardini A, Savin M, Menditto E, Lehane E, Laosa O, et al. Interventional tools to improve medication adherence: review of literature. Patient Prefer Adherence 2015;9:1303-1314 [FREE Full text] [doi: 10.2147/PPA.S87551] [Medline: 26396502]
33. Farooqi M, Carter C, Patel T. Electronic medication adherence technologies’ classification to guide use in older adults: Pharmacy practice research abstracts: Canadian Pharmacists Conference 2017. Can Pharm J 2017;150(4):S1-72 [FREE Full text] [doi: 10.1177/1715163517719855]
34. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mhealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
35. Daniels J, Fels S, Kushniruk A, Lim J, Ansermino JM. A framework for evaluating usability of clinical monitoring technology. J Clin Monit Comput 2007 Oct;21(5):323-330. [doi: 10.1007/s10877-007-9091-y] [Medline: 17701381]
36. Wildenbos G, Peute L, Jaspers M. A framework for evaluating mhealth tools for older patients on usability. Stud Health Technol Inform 2015;210:783-787. [doi: 10.3233/978-1-61499-512-8-783] [Medline: 25991261]
37. Reeder B, Demiris G, Marek KD. Older adults' satisfaction with a medication dispensing device in home care. Inform Health Soc Care 2013 Sep;38(3):211-222 [FREE Full text] [doi: 10.3109/17538157.2012.741084] [Medline: 23323721]
38. Tseng M, Wu H. A cloud medication safety support system using QR code and web services for elderly outpatients. Technol Health Care 2014;22(1):99-113. [doi: 10.3233/THC-140778] [Medline: 24561883]
39. Ligons FM, Romagnoli KM, Browell S, Hochheiser HS, Handler SM. Assessing the usability of a telemedicine-based medication delivery unit for older adults through inspection methods. AMIA Annu Symp Proc 2011;2011:795-804 [FREE Full text] [Medline: 22195137]
40. Ligons FM, Mello-Thoms C, Handler SM, Romagnoli KM, Hochheiser H. Assessing the impact of cognitive impairment on the usability of an electronic medication delivery unit in an assisted living population. Int J Med Inform 2014 Nov;83(11):841-848 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.07.004] [Medline: 25153770]
41. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
42. Stuck RE, Chong AW, Mitzner TL, Rogers WA. Medication management apps: usable by older adults? Proc Hum Factors Ergon Soc Annu Meet 2017 Sep;61(1):1141-1144 [FREE Full text] [doi: 10.1177/1541931213601769] [Medline: 29158662]
43. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
44. Brook J. SUS-a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. London, UK: CRC Press; 1996.
45. Bangor A, Staff T, Kortum P, Miller J, Staff T. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]
46. Hart SG, Staveland LE. Development of NASA-TLX (task load ondex): results of empirical and theoretical approach. Adv Psychol 1988;52:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
47. Grier RA. How high is high? A meta-analysis of NASA-TLX global workload scores. Proc Hum Factors Ergon Soc 2015;59(1):1727-1731. [doi: 10.1177/1541931215591373]
48. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
49. Sauro J, Lewis J. Quantifying the User Experience: Practical Statistics for User Research. Second Edition. San Francisco, California: Morgan Kaufmann; 2016.

Abbreviations

ANOVA: analysis of variance
NASA-TLX: NASA-task load index
SUS: system usability scale

Edited by G Eysenbach; submitted 01.02.20; peer-reviewed by B McKinstry, KZ Yap; comments to author 10.03.20; revised version received 25.03.20; accepted 28.03.20; published 02.06.20.

Please cite as:
Patel T, Ivo J, Faisal S, McDougall A, Carducci J, Pritchard S, Chang F
A Prospective Study of Usability and Workload of Electronic Medication Adherence Products by Older Adults, Caregivers, and Health Care Providers
J Med Internet Res 2020;22(6):e18073
URL: https://www.jmir.org/2020/6/e18073
doi:10.2196/18073
PMID:32348292

Original Paper

Medical Mobile App Classification Using the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies: Interrater Reliability Study

Khine Nwe[1], BSc, MSc; Mark Erik Larsen[2], DPhil; Natalie Nelissen[3], MSc, PhD; David Chi-Wai Wong[4,5], DPhil, MEng

[1] Leeds Institute of Health Sciences, University of Leeds, Leeds, United Kingdom
[2] Black Dog Institute, University of New South Wales, Sydney, Australia
[3] Leeds Institute of Data Analytics, University of Leeds, Leeds, United Kingdom
[4] Centre for Health Informatics, University of Manchester, Manchester, United Kingdom
[5] Department of Computer Science, University of Manchester, Manchester, United Kingdom

Corresponding Author:
David Chi-Wai Wong, DPhil, MEng
Centre for Health Informatics
University of Manchester
Vaughan House
Portsmouth Street
Manchester, M13 9GB
United Kingdom
Phone: 44 1613069280
Email: david.wong@manchester.ac.uk

Abstract

Background: Clinical governance of medical mobile apps is challenging, and there is currently no standard method for assessing the quality of such apps. In 2018, the National Institute for Health and Care Excellence (NICE) developed a framework for assessing the required level of evidence for digital health technologies (DHTs), as determined by their clinical function. The framework can potentially be used to assess mobile apps, which are a subset of DHTs. To be used reliably in this context, the framework must allow unambiguous classification of an app's clinical function.

Objective: The objective of this study was to determine whether mobile health apps could be reliably classified using the NICE evidence standards framework for DHTs.

Methods: We manually extracted app titles, screenshots, and content descriptions for all apps listed on the National Health Service (NHS) Apps Library website on July 12, 2019; none of the apps were downloaded. Using this information, 2 mobile health (mHealth) researchers independently classified each app to one of the 4 functional tiers (ie, 1, 2, 3a, and 3b) described in the NICE digital technologies evaluation framework. Coders also answered contextual questions from the framework to identify whether apps were deemed to be higher risk. Agreement between coders was assessed using Cohen κ statistic.

Results: In total, we assessed 76 apps from the NHS Apps Library. There was classification agreement for 42 apps. Of these, 0 apps were unanimously classified into Tier 1; 24, into Tier 2; 15, into Tier 3a; and 3, into Tier 3b. There was disagreement between coders in 34/76 cases (45%); interrater agreement was poor (Cohen κ=0.32, 95% CI 0.16-0.47). Further investigation of disagreements highlighted 5 main explanatory themes: apps that did not correspond to any tier, apps that corresponded to multiple tiers, ambiguous tier descriptions, ambiguous app descriptions, and coder error.

Conclusions: The current iteration of the NICE evidence standards framework for DHTs did not allow mHealth researchers to consistently and unambiguously classify digital health mobile apps listed on the NHS app library according to their functional tier.

(J Med Internet Res 2020;22(6):e17457) doi:10.2196/17457

KEYWORDS

mHealth; telehealth; evaluation; evidence; interrater; NHS Apps Library; NICE

Introduction

Mobile apps for use in health care have been proposed in a variety of settings, including telehealth for disease management and monitoring, diagnosis and triage, and medication prescription and reminders [1-3]. To date, the evidence for their effectiveness is varied [4-6] and, in general, published evidence is extremely sparse [7].

Thorough governance of medical apps is challenging [8]. There is growing recognition that the required level of evidence for medical apps ought to differ depending on app function. For instance, Wyatt [9] suggested that apps be categorized as low, medium, or high risk for purposes of evaluation. The UK National Institute for Health and Care Excellence (NICE) further addressed this issue through their recent evidence standards framework for evaluating digital health technologies (DHTs) [10]. In this framework, digital technologies are categorized into one of 3 tier levels (with Tier 3 split into Tier 3a and 3b) according to their intended function. An additional set of questions assesses higher-risk technologies, complementing the tier levels.

The framework also provides minimum and best-practice recommendations on the associated standard of evidence required for each tier. For instance, Tier 3a technologies should be supported, at a minimum, by relevant *high-quality observational* or *quasi-experimental studies.* Best practice would include a *high-quality intervention study*. The recommended evidence standards are cumulative, such that Tier 3 digital technologies should also meet the criteria for Tier 2 and Tier 1 technologies. For Tier 2, this includes, for instance, having evidence to show that any health information provided is valid, accurate, up-to-date, regularly audited, and sufficiently comprehensive. At the time of writing, the framework is in its second iteration.

Although other frameworks exist and have been used for classifying mobile apps [11], we solely examine the NICE framework here. This framework is of particular importance as its development was commissioned by the National Health Service (NHS) England and is therefore likely to become an influential standard. In addition, its recency means that it has thus far received little external validation.

This study aims to evaluate the appropriateness and potential limitations of the functional classification guidance within the NICE framework as applied to trusted and safe mobile health (mHealth) apps. We will do this by assessing interrater agreement of functional tier classification for all apps curated on the NHS Apps Library. By examining cases in which reviewers disagreed, we will highlight ambiguities in the current classification guidance and discuss potential improvements.

Methods

Search Strategy and Data Extraction

All apps available on the NHS Apps Library [12] on July 12, 2019, were included in the analysis. For each app, the app title, screenshots, and description were extracted manually from the NHS Apps Library website; none of the apps were downloaded. No apps were excluded.

Classification

Two coders independently classified all apps according to functional classification. The coders were a clinician with formal postgraduate training in health informatics (KN) and an academic with research expertise in mHealth (MEL).

For each app, we recorded the main features as described on the NHS Apps Library, including any available screenshots. The coders assigned each app to a functional tier and noted whether the app should be considered for risk adjustment based on clinical context. Abridged information about each tier and criteria for determining risk-adjusted apps are shown in Tables 1 and 2. To guide our classification decisions, we used the (unabridged) evidence standards framework alongside the associated user guide [13]. If an app could be assigned into multiple tiers, it was assigned to the highest relevant tier, as per the framework guidance. If apps met the criteria for both Tier 3a and Tier 3b, they were assigned to 3b.

Table 1. Abridged definitions of digital health technology tiers from the National Institute for Health and Care Excellence (NICE) evidence standards evaluation framework [10].

Tier and functional classification	Description
1	
System service	Improves system efficiency. Unlikely to have direct and measurable individual patient outcomes.
2	
Information	Provides information and resources to patients or the public.
Simple monitoring	Allows users to record health parameters to create health diaries.
Communicate	Allows two-way communication between users and professionals, carers, third-party organizations, or peers.
3a	
Preventative behavior change	Designed to change user behavior related to health issues with, for example, smoking, eating, alcohol, sexual health, sleeping, and exercise.
Self-manage	Aims to help people with a diagnosed condition to manage their health.
3b	
Treat	Provides treatment for a diagnosed condition (such as cognitive behavioral therapy for anxiety), or guides treatment decisions.
Active monitoring	Automatically records information and transmits the data to a professional, carer, or third-party organization, without any input from the user, to inform clinical management decisions.
Calculate	Tools that perform clinical calculations that are likely to affect clinical care decisions.
Diagnose	Uses data to diagnose a condition in a patient, or to guide a diagnostic decision made by a health care professional.

Table 2. Abridged contextual questions to help identify higher-risk digital health technologies (DHTs), abridged from [10].

Question	Risk adjustment
1. Are the intended users of the DHT considered to be in a potentially vulnerable group such as children or at-risk adults?	The National Health Service (NHS) England defines an at-risk adult as an adult "who may be in need of community care services by reason of mental or other disability, age or illness; and who is or may be unable to take care of him or herself, or unable to protect him or herself against significant harm or exploitation." If the DHT is intended to be used by people considered to be in a potentially vulnerable group, then a higher level of evidence may be needed, or relevant expert opinion on whether the needs of the users are being appropriately addressed.
2. How serious could the consequences be to the user if the DHT failed to perform as described?	A higher level of potential harm may indicate that the best-practice evidence standards should be used.
3. Is the DHT intended to be used with regular support from a suitably qualified and experienced health or social care professional?	DHTs that are intended to be used with support (ie, with regular support or guidance from a suitably qualified and experienced health or social care professional) could be considered to have lower risk than DHTs that are intended to be used by the patients on their own. This contextual question may require careful interpretation depending on the individual DHT as the involvement of a clinician may in itself indicate that the DHT presents a specific risk.
4. Does the DHT include machine learning algorithms or artificial intelligence?	Refer to the code of conduct for data-driven health and care technology for additional considerations when assessing DHTs that use artificial intelligence or machine learning.
5. Is the financial or organizational risk of the DHT expected to be very high?	DHTs with very high financial risk should be assessed using the best-practice standards to provide surety that the DHT represents good value. High organizational risks may include situations in which implementing the DHT would need complex changes in working practice or care pathways.

Classification

We used descriptive statistics to summarize the membership of each tier and the differences in classification between reviewers. We reported Cohen κ as an overall measurement of interrater agreement [14]. In supplementary analysis, we reported interrater agreement for a subset of apps that had previously been classified by another independent team and published in the framework user guide. All analyses were conducted using MATLAB (version 18.1; MathWorks) [15].

We examined any apps with discrepancies in functional tier classification in greater detail. Using the publicly available descriptions of each app, we identified common themes that may have led to differences in classification.

Results

Interrater Agreement Results

In total, we assessed 76 apps from the NHS Apps Library. Overall, 0 apps were unanimously classified into Tier 1; 24/76 (32%), into Tier 2; 15/76 (20%), into Tier 3a; and 3/76 (4%), into Tier 3b. Full classification details for each app are presented in Multimedia Appendix 1.

Table 3 shows the interrater agreement for each tier of app. There was disagreement in 34/76 cases (45%) and Cohen κ was 0.32 (95% CI 0.16-0.47), which is commonly considered to correspond to poor agreement [12]. Of the 34 apps for which functional classification differed, 13 were due to discrepancies between apps classified in Tier 3a or 3b. The next largest group of discrepancies was with apps classified between Tiers 2 and 3a (n=11). Analysis of interrater agreement for a subset of apps previously reviewed by a third independent group is presented in Multimedia Appendix 2 (coder 1: κ=0.48, 95% CI 0.31-0.66; coder 2: κ=0.62, 95% CI 0.44-0.80). The apps for which the independent coders disagreed are also listed in Multimedia Appendices 1 and 2.

Coders also assessed on whether an app should be considered for *risk adjustment* or not based on a set on contextual questions (Table 2). Of these, Questions 2, 4, and 5 were not possible to answer without wider knowledge of the source code (for Question 4) or how the app interacted with the wider health care system. A total of 9 apps were unanimously considered to warrant risk adjustment, and 63 for no adjustment. There were discrepancies for 4 apps, and overall agreement may be considered good (κ=0.79; 95% CI 0.59-0.99).

Table 3. Number of apps coded into each functional tier classification, for both coders.

Classification: coder 1	Classification: coder 2			
	1	2	3a	3b
1	0	6	0	0
2	1	24	5	3
3a	0	6	15	1
3b	0	0	12	3

Analysis of Coder Discrepancies

On further collaborative review, differences in tier classification could be attributed to 5 causes: (1) app function not listed within framework, (2) app function corresponded to multiple tiers, (3) ambiguous tier description, (4) ambiguous app description, and (5) human error.

App Function Not Listed Within Framework

This occurred when the intended function of the app, as described on the NHS Apps Library, did not correspond to any of the functions listed within the NICE framework. One example of this was *MyChoicePad*, an app that is designed to assist nonverbal communication via symbols and signs from the Makaton language. Although one coder categorized the app and function as Tier 2: Communicate, the app does not facilitate two-way communication, so it does not strictly meet the associated criteria. Similarly, the other reviewers categorized the app into Tier 1: System service, but it is not entirely clear whether the app is designed to improve system efficiency, or even what the system is in this case.

App Function Corresponded to Multiple Tiers

This occurred when a single function of an app corresponded to more than 1 tier within the NICE framework. In particular, we noted ambiguity around mental health apps. For example, mindfulness or principles from cognitive behavioral therapy may be classified as a self-management strategy to reduce feelings of anxiety (3a), or as a treatment for anxiety disorders (3b). In this situation, the tier classification depended on the clinical use case, rather than the app function, which was the same in both cases. We also noted ambiguity between lifestyle/well-being versus diagnosed conditions. For example, an app that had a function to provide advice on stress or anxiety reduction might be classified into Tier 2 if it provides generic information, but into 3a if *stress* were part of a diagnosed condition.

Ambiguous Tier Description

Some terms used to describe the tiers within the NICE framework were not clearly defined. One specific instance of ambiguity between Tiers 3a and 3b occurred for the *engage warfarin self-care* app. The app allows warfarin test results to be *actively reviewed* by a clinician, but it is unclear whether this constitutes Tier 3b's *active monitoring* as results must be manually transmitted by the user, or *simple monitoring* which allows users to record their health parameters.

Ambiguous App Description

In some instances, the information provided on the NHS Apps Library was insufficient to definitively categorize an app. For example, screenshots on the NHS Apps Library for the *NHS App* show that users can *check symptoms*, but this functionality was not mentioned elsewhere. The main description for the app stated that users can "find reliable NHS information on hundreds of conditions and treatments, and get immediate advice." If the app merely catalogs searchable information on symptoms, it should be classed as Tier 2; however, if it provides a symptom checker algorithm that requires user input, and outputs relevant immediate advice, it is Tier 3b.

Human Error

This occurred when the reviewer failed to identify a relevant piece of information that would have influenced their tier categorization decision. This most commonly occurred when an app had multiple functions belonging to multiple functional tiers. Typically, the main stated function of the app belonged to the lower tier. For instance, the *Healthera* app is primarily designed to manage prescription management, a Tier 1 function. However, it also allows users to contact their pharmacist for clinical advice via the app, a Tier 2 function. Human errors accounted for 6 apps. In conjunction with the primary result, we can estimate an upper bound on the level of disagreement as 28/76 (37%; κ=0.44, 95% CI 0.30-0.60), when human error is removed.

Discussion

Principal Results

Our results show that, from their publicly available descriptions, only 42 of the 76 apps collated in the NHS Apps Library could be classified into a functional tier consistently by informed individuals. Interrater agreement between reviewers was poor (κ=0.32).

Of the 34 apps where there was disagreement, there was a subset (28/34) for which the tier could not be agreed, even after consultation (ie, excluding human error). In these cases, disagreement was attributed to four broad categories: *App function not listed within framework*, *App function corresponded to multiple tiers*, *Ambiguous tier description*, and *Ambiguous app description*.

There was good agreement in assessment of *higher risk* using the NICE framework's contextual questions (κ=0.79). Despite this, we noted ambiguity in the risk adjustment questions. One example of ambiguity occurs for the assessment question: "Does the DHT include machine learning algorithms or artificial intelligence?" If yes, framework users are advised to refer to the Code of Conduct for data-driven health and care technology [16].

Whilst the framework defines both artificial intelligence and machine learning, the provided definitions are ambiguous. For instance, it states that "Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed." This definition would include the vast majority of quantitative methods. For instance, linear regression is fully defined by the slope and offset parameters, as learned directly from a data set.

The Code of Conduct further provides an external link to the definition from the AHSN Network AI Initiative [17]. No definition is provided here, instead readers are informed that "there is no single, universally agreed definition of AI." The absence of any clear definition means that this risk adjustment question cannot be answered objectively.

Poor interrater agreement of tier classification may be attributed to two potential causes. First, the publicly available information on the NHS App Library may be insufficient to determine the functional tier. If true, this would motivate tighter regulation of how apps are described on the Library to ensure that the intended medical condition and patient or user group is clear. Second, the DHT framework is not specific enough to classify some types of apps. Based on our thematic analysis, in which we showed examples of misclassified apps due to ambiguity in the framework, we contend that this second reason contributes significantly to the overall level of disagreement.

A framework with inadequate specificity has implications for both developers and regulators. For developers wishing to bring products to market as soon as possible, opportunity for misclassification due to ambiguous tier criteria may result in more classifications to lower tiers, where the minimum standard of evidence is not as stringent. In particular, Tier 2 technologies require only information that would be commonly audited in standard software development, whereas Tier 3a and Tier 3b technologies specifically require formal studies that would likely require additional time and financial resource. This may lead to situations in which apps are regulated to a level of lower scrutiny than they ought to be, given their function.

In addition to the main result, we observed that some apps were categorized consistently by reviewers, but the minimum suggested level of evidence did not seem to align with the level of potential clinical risk. One example of this was *Cypher*, which was classified in Tier 2 as it facilitates communication. Whereas other Tier 2 apps allowed communication with health professionals, this app facilitates anonymous communication with other users to allow "anyone who want to share their thoughts". (We further note that as of October 17, 2019, Cypher App is not available for download on the either the Google Play store or Apple App Store and that the developer website is nonfunctional. Persistence of apps is a known problem in digital health [18].) The evidence framework directly addresses this use case by requiring peer-support apps to show evidence of appropriate safeguarding. However, internet communities, forums, and chat rooms [19,20] with similar peer-support functions have been shown to have potential for negative patient impact. Given that the potential harm aligns more closely with apps in Tier 3a or 3b, a higher standard of evidence may be appropriate for peer-support apps, despite the similarity in technical function to other communication apps.

Within mHealth more widely, we can select examples in which the dissonance between functional tier and required evidence is even greater. For instance, consider an app that calculates BMI by requiring the user to type height and weight. This may be categorized into Tier 3b, as a tool that performs a clinical calculation that can affect clinical care decisions. The associated minimum standard of evidence in the NICE framework requires a high-quality intervention study—a level of scrutiny that ought not to be required for simple and well-established calculations. This does not preclude the need for careful technical evaluation; indeed, Huckvale et al [21] have demonstrated how even simple clinical calculations in diabetes apps are often calculated or displayed incorrectly.

More broadly, this indicates that clinical risk and technical risk are not necessarily the same. One possible option to enable better classification and more specific evaluation guidance may be to categorize apps by technical complexity as well as clinical function. The idea of separating clinical and technical evaluation has been raised previously. Lewis and Wyatt [22] suggest evaluation could be based on the probability and severity of clinical harm, the complexity of the app, and additional contextual factors [22]. Others have commented that evidence of clinical effectiveness for software should include indication of safety, and that this must include formal technical evaluation [23].

Limitations

The NICE framework is designed for DHTs that are commissioned by the health system. In this case, we examined apps on the NHS Apps Library. According to the library's HTML description metadata, these are "digital tools that have been assessed by the NHS as clinically safe and secure to use" [12]. However, in most instances, the apps had not been specifically commissioned. Although the relevance of the framework for these apps may vary, multiple apps in the library (eg, GDm-Health, Health Help Now) have already been used as part of routine clinical care.

The primary limitation of work is that our classifications relied on the information presented by the NHS App Library; additionally, none of the apps were downloaded. As all products published in the Library met a set of internal standards, we believed, a priori, that written descriptions and screenshots should be sufficient to enable clear identification of all key functions (as this is key information for informing consumer app selection). Our results showed that this was not the case. Downloading each app would provide more comprehensive understanding of the key functions and may increase interrater agreement on app tier. We therefore recommend that future studies download and engage directly with the contents of mobile apps.

Our study compares only two sets of raters, so the results might be unduly influenced by the poor performance of a single individual. We partly addressed this by comparing interrater reliability with a third set of reviewers (see Multimedia Appendices 1 and 2), for which similar results were obtained. These were not included in the main manuscript as reviews were conducted under different conditions; we do not know if apps were downloaded, nor the time at which their review took place.

Finally, the NICE framework establishes a functional tier and provides guidance on required levels of evidence at each tier. In this work, we only examined the consistency of tier classification, and did not address whether apps within a tier met the evidence standards.

Conclusions

The NICE evidence standards framework for evaluating digital technologies is a significant and timely step toward establishing appropriate levels of evidence for DHTs. Despite this, we have demonstrated that the current iteration of the framework did not allow mHealth researchers to consistently and unambiguously group a set of digital health mobile apps according to their functional tier. In users with limited experience of mHealth evaluation (eg, app developers), we postulate that this ambiguity may lead to higher levels of misclassification. One potential improvement would be to classify DHTs by their technical complexity in addition to clinical function.

Acknowledgments

KN was supported by a Chevening Scholarship.

Authors' Contributions

All authors contributed to the design, analysis, and final manuscript writing. KN undertook the initial data extraction.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Raw data used for analysis. Independent data were extracted from NICE case studies.

[DOCX File , 20 KB - jmir_v22i6e17457_app1.docx]

Multimedia Appendix 2

Inter-rater reliability crosstables for coders 1 and 2 vs the case study examples provided with the evidence standards framework.

[DOCX File , 14 KB - jmir_v22i6e17457_app2.docx]

References

1. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
2. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015;351:h3480 [FREE Full text] [Medline: 26157077]

3. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
4. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
5. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Dec 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]
6. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
7. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
8. Magrabi F, Habli I, Sujan M, Wong D, Thimbleby H, Baker M, et al. Why is it so difficult to govern mobile apps in healthcare? BMJ Health Care Inform 2019 Nov;26(1):e100006 [FREE Full text] [doi: 10.1136/bmjhci-2019-100006] [Medline: 31744843]
9. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 03;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
10. Evidence Standards Framework for Digital Health Technologies. London: National Institute for Health and Care Excellence; 2019 Mar. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2019-10-17]
11. Olla P, Shimskey C. mHealth taxonomy: a literature survey of mobile health applications. Health Technol 2015 Jan 30;4(4):299-308. [doi: 10.1007/s12553-014-0093-8]
12. NHS. NHS Apps Library. London: NHS URL: https://www.nhs.uk/apps-library/ [accessed 2019-10-17]
13. National Institute for Health and Care Excellence. Evidence Standards Framework for Digital Health Technologies: User Guide. 2019 Mar. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/user-guide.pdf [accessed 2019-10-17]
14. Fleiss J, Levin B, Paik M. The measurement of interrater agreement. In: Statistical Methods for Rates and Proportions. 3rd edition. New York: John Wiley & Sons Inc; 2003.
15. MathWorks. MATLAB and Statistics Toolbox Release. Natick, MA: MathWorks; 2018.
16. Department of Health & Social Care. Code of Conduct for Data-Driven Health and Care Technology (Updated 18 July 2019). London: Department of Health & Social Care URL: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology [accessed 2019-10-17]
17. The AHSN Network. AI in Health and Care: What Do We Mean by AI in Health and Care. URL: http://ai.ahsnnetwork.com/about/ai-in-health-and-care/ [accessed 2019-10-17]
18. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016 Aug 09;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
19. Beebe TJ, Asche SE, Harrison PA, Quinlan KB. Heightened vulnerability and increased risk-taking among adolescent chat room users: results from a statewide school survey. J Adolesc Health 2004 Aug;35(2):116-123. [doi: 10.1016/j.jadohealth.2003.09.012] [Medline: 15261640]
20. Daine K, Hawton K, Singaravelu V, Stewart A, Simkin S, Montgomery P. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. PLoS One 2013;8(10):e77555 [FREE Full text] [doi: 10.1371/journal.pone.0077555] [Medline: 24204868]
21. Huckvale K, Adomaviciute S, Prieto JT, Leow MKS, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015 May 06;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
22. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
23. Thimbleby H, Lewis A, Williams J. Making healthcare safer by understanding, designing and buying better IT. Clin Med (Lond) 2015 Jun;15(3):258-262 [FREE Full text] [doi: 10.7861/clinmedicine.15-3-258] [Medline: 26031976]

Abbreviations

DHT: digital health technologies
NHS: National Health Service
NICE: National Institute for Clinical Excellence

Edited by G Eysenbach; submitted 16.12.19; peer-reviewed by L Valentine, F Holl; comments to author 13.02.20; revised version received 25.02.20; accepted 29.02.20; published 05.06.20.

Please cite as:
Nwe K, Larsen ME, Nelissen N, Wong DCW
Medical Mobile App Classification Using the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies: Interrater Reliability Study
J Med Internet Res 2020;22(6):e17457
URL: http://www.jmir.org/2020/6/e17457/
doi:10.2196/17457
PMID:32501271

Original Paper

Reliability of Smartphone for Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores in Acute Ischemic Stroke Patients: Diagnostic Test Accuracy Study

Kenichiro Sakai[1], MD, PhD; Teppei Komatsu[1], MD, PhD; Yasuyuki Iguchi[1], MD, PhD; Hiroyuki Takao[2], MD, PhD; Toshihiro Ishibashi[2], MD, PhD; Yuichi Murayama[2], MD, PhD

[1]Department of Neurology, The Jikei University School of Medicine, Tokyo, Japan
[2]Department of Neurosurgery, The Jikei University School of Medicine, Tokyo, Japan

Corresponding Author:
Kenichiro Sakai, MD, PhD
Department of Neurology
The Jikei University School of Medicine
Nishishinbashi 3-25-8
Tokyo, 1050003
Japan
Phone: 81 3 3433 1111
Email: k.sakai.0127@gmail.com

Abstract

Background: High-quality neuroimages can be viewed using a medical app installed on a smartphone. Although interdevice agreement between smartphone and desktop PC monitor was found to be favorable for evaluating computed tomography images, there are no interdevice agreement data for diffusion-weighted imaging (DWI).

Objective: The aim of our study was to compare DWI interpretation using the Join smartphone app with that using a desktop PC monitor, in terms of interdevice and interrater agreement and elapsed interpretation time.

Methods: The ischemic change in the DWI of consecutive patients with acute stroke in the middle cerebral artery territory was graded by 2 vascular neurologists using the Join smartphone app and a desktop PC monitor. The vascular neurologists were blinded to all patient information. Each image was categorized as either Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores (DWI-ASPECTS) ≥7 or DWI-ASPECTS <7 according to the Japanese Society for Neuroendovascular Therapy. We analyzed interdevice agreement and interrater agreement with respect to DWI-ASPECTS. Elapsed interpretation time was compared between DWI-ASPECTS evaluated by the Join smartphone app and a desktop PC monitor.

Results: We analyzed the images of 111 patients (66% male; median age=69 years; median National Institutes of Health Stroke Scale score on admission=4). Interdevice agreement regarding DWI-ASPECTS between the smartphone and the desktop PC monitor was favorable (vascular neurologist 1: κ=0.777, *P*<.001, vascular neurologist 2: κ=0.787, *P*<.001). Interrater agreement was also satisfactory for the smartphone (κ=0.710, *P*<.001) and the desktop PC monitor (κ=0.663, *P*<.001). Median elapsed interpretation time was similar between the smartphone and the desktop PC monitor (vascular neurologist 1: 1.7 min vs 1.6 min; *P*=.64); vascular neurologist 2: 2.4 min vs 2.0 min; *P*=.14).

Conclusions: The use of a smartphone app enables vascular neurologists to estimate DWI-ASPECTS accurately and rapidly. The Join medical smartphone app shows great promise in the management of acute stroke.

(J Med Internet Res 2020;22(6):e15893) doi:10.2196/15893

KEYWORDS

smartphone app; DWI; ASPECTS

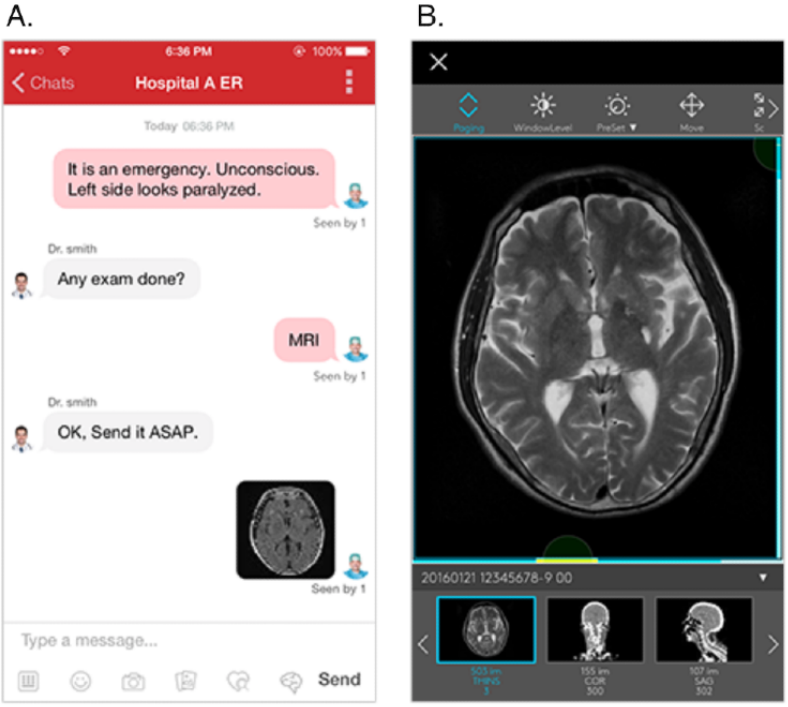
Introduction

In 1996, the first-line treatment for acute-onset ischemic stroke was intravenous thrombolysis using recombinant tissue plasminogen activator (IV rtPA) therapy, which was effective only within 3 hours after the onset of symptoms. Since then, the emergency medical systems for acute stroke patients have improved dramatically. The publication of the ECASS III (European Acute Stroke Study III) [1], DAWN (Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo) [2], and DEFUSE 3 (Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution) [3] trials and advances such as those reported from the WAKE-UP (Wake-Up Stroke) trial [4] and new thrombolytic agents [5] have expanded the therapeutic time window and increased the number of candidates suitable for IV rtPA and mechanical thrombectomy. Accordingly, the decision-making process for thrombolysis requires timelier, more accurate, and more professional neurological assessment (including neuroimaging) to be made by a stroke specialist. The sharing of clinical and neuroimaging information will become increasingly important in decision making for IV rtPA and mechanical thrombectomy, particularly at comprehensive stroke centers. There is an urgent need to build a more convenient and faster communication system for sharing this information among the stroke team, which comprises vascular neurologists, on-call physicians, residents, and emergency, operating room, and paramedical staff.

We used the Join medical smartphone app to build a seamless communication system for the stroke team. The app enables the team to share texts, neuroimaging, photos, and videos with high security (Figure 1A). Immediately after neuroimaging of a stroke patient, the images are sent from the hospital server to all smartphones that have the app installed and are signed in as members of the stroke team. The images can be enlarged and evaluations recorded with a simple touch sequence on the smartphone screen (Figure 1B). Before this communication system can be used in the newly extended therapeutic window in acute stroke, it is necessary to confirm interdevice agreement and interrater agreement with regard to assessment of the neuroimages.

At our stroke center, the initial imaging examination for patients with acute ischemic stroke is magnetic resonance imaging (MRI) rather than computed tomography (CT), for the following reasons: (1) hyperacute ischemic stroke is easily diagnosed on diffusion-weighted imaging (DWI) on MRI and (2) cerebral artery occlusion can be assessed on magnetic resonance angiography without the use of contrast agents. Although interdevice agreement between the smartphone and a desktop PC monitor was found to be favorable for evaluating CT images [6], there are no interdevice agreement data for DWI. The aim of our study was to compare DWI interpretation using the Join smartphone app with that using a desktop PC monitor, in terms of interdevice and interrater agreement and elapsed interpretation time.

Figure 1. A. The Join smartphone app utilizes the easy-to-use interface of the social networking communication environment. B. Communication with picture archiving and communication system and other intrahospital systems enable text and medical images hosted on a cloud server to be shared in a group chat. The Join smartphone app displays diagnostic medical images, such as MRI and CT, and enables app users to edit, comment on, and draw a shape. CT: computed tomography. MRI: magnetic resonance imaging.



Today 06:36 PM

It is an emergency. Unconscious. Left side looks paralyzed.

Seen by 1

Dr. smith

Any exam done?

MRI

Seen by 1

Dr. smith

OK, Send it ASAP.

Seen by 1

20160121 12345678-9 00

Methods

Patient Characteristics

From January 2016 to September 2017, we enrolled 111 patients with acute ischemic stroke in the middle cerebral artery territory, diagnosed within 24 hours of onset. DWI on MRI was performed on all patients, and the following clinical information was recorded: cardiovascular risk factors (hypertension, diabetes mellitus, dyslipidemia, and smoking status) and atrial fibrillation. Stroke severity on admission was assessed using the National Institutes of Health Stroke Scale (NIHSS) score. Stroke subtype was categorized into four groups: small-vessel occlusion, large-artery atherosclerosis, cardioembolism, or other.

Join Smartphone App

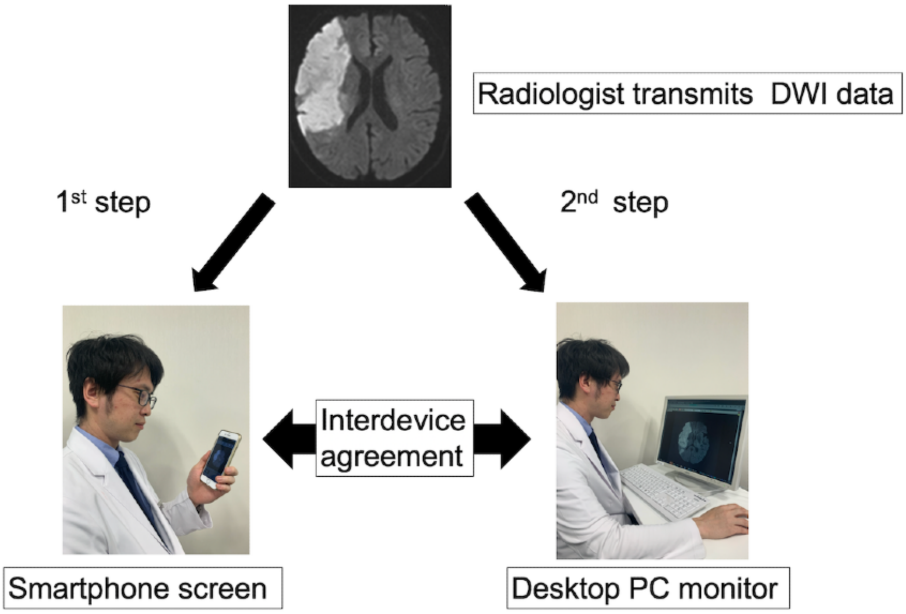
We evaluated the Join smartphone app (Allm Inc.), which was developed for use as a telemedicine app for health care professionals. The Join smartphone app leverages the easy-to-use interface of the social networking communication environment, such as SMS text messaging (Figure 1A) and, importantly, enables the stroke team to immediately share medical information such as diagnostic images (CT, MRI, and ultrasonography) and electrocardiograms as well as the results of blood tests (Figure 1B). Information sharing begins as soon as the emergency department is informed of an incoming potential stroke patient and continues as relevant personnel are called; the initial diagnostic and therapeutic orders are prepared, and senior staff are consulted if necessary. Following the acquisition of imaging studies, the images and radiological reports are shared. Additional details of the patient evaluation (including digital video recordings of clinical signs) can be requested and sent to the senior consulting staff. Following discussion among the team, the final management decisions are made before the patient is admitted to the stroke care unit. For purposes of security, no information is stored on any smartphone, and the app displays only the medical information and images that are streamed from the cloud server. On completion of the communication session, no discussion related to the patient remains on the smartphone.

Imaging

DWI was performed with a 1.5 T MRI unit (MAGNETOM Avanto/MAGNETOM Symphony, Siemens) using the following sequence: repetition time/echo time=2700/90 ms, section thickness=5 mm, section gap=1.5 mm, matrix=128x128, field of view=21 cm. Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores (DWI-ASPECTS) were defined using a scoring template comprising 2 axial DWI slices with markers for 10 anatomical regions [7,8]. We checked the entire sequence of DWI slices to calculate the score. Each patient was categorized as DWI-ASPECTS ≥7 or DWI-ASPECTS <7, according to the statement of the Japanese Society for Neuroendovascular Therapy for patients undergoing mechanical thrombectomy.

The study protocol was as follows (Figure 2). Two vascular neurologists (KS and TK) installed the Join smartphone app on their smartphones. After confirming operation of the app, a radiologist transmitted the DWI data of all patients to the vascular neurologists, in random order. As the first step, the vascular neurologists received the DWI data on their smartphones and independently scored DWI-ASPECTS for all patients using the Join smartphone app (JOIN-ASPECTS). The time for transfer of the DWI data was a few seconds. As the second step, vascular neurologists interpreted all DWI data on a desktop PC monitor and individually scored DWI-ASPECTS (PC-ASPECTS) a few days later. The vascular neurologists were blinded to the background and clinical information of all patients. We recorded the time taken for each vascular neurologist to complete DWI scoring for JOIN-ASPECTS and PC-ASPECTS**.** This study conformed to the ethical principles established in the Declaration of Helsinki, and the Institutional Review Board at the Jikei University School of Medicine approved the study protocol (no. 8813).

Figure 2. Study protocol. First, the vascular neurologists evaluated DWI-ASPECTS on a smartphone screen. Second, they evaluated DWI-ASPECTS on a desktop PC monitor. Interdevice agreement and interrater agreement were calculated for the same and for different devices. DWI-ASPECTS: Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores.



Statistical Analysis

To calculate interdevice (smartphone and desktop PC monitor) agreement, we prepared scatter diagrams for JOIN-ASPECTS and PC-ASPECTS. Scatter diagrams were created for all patients, including patients with major arterial occlusion and those without major arterial occlusion. Kappa statistics were used to calculate interdevice and interrater (vascular neurologists KS and TK) agreement.

We defined a component of interrater agreement among the vascular neurologists as the agreement rate of DWI-ASPECTS in each patient, for a DWI cutoff of ≥7 or <7. We evaluated interrater agreement between the vascular neurologists for the smartphone, the desktop PC monitor, and then for both the smartphone and the desktop PC monitor. Interrater agreement was assessed using 2x2 cross-tabulation.

Kappa scores were rated as follows: <0.20, poor agreement; 0.21-0.40, fair agreement; 0.41-0.60, moderate agreement; 0.61-0.80, favorable agreement; and 0.81-1.0, almost perfect agreement. *P*<.05 was considered significant. All statistical analyses were performed using SPSS for Windows, version 22.0 (IBM Corp.).

Results

We enrolled 111 patients (66% male; median age, 69 years; median NIHSS score on admission, 4). Table 1 lists the patients' clinical characteristics. Median DWI-ASPECTS was 9 (6-10), and 46 (41%) patients had major artery occlusion. The median elapsed time between symptom onset and DWI imaging was 270 min. Interdevice agreement between the smartphone and the desktop computer monitor was favorable (KS: κ=0.777, *P*<.001; TK: κ=0.787, *P*<.001) for all patients (Figures 3A and 3B). Interdevice agreement was also favorable for patients with and those without major arterial occlusion (Figures 4A, 4B, 5A, and 5B). The median elapsed interpretation times (from receiving the image to finishing interpretation) were similar for the Join smartphone app and the desktop PC monitor (KS: 1.7 min vs 1.6 min, *P*=.64; TK: 2.4 min vs 2.0 min, *P*=.14). Interrater agreement between the 2 vascular neurologists was favorable for the Join smartphone app (κ=0.710, *P*<.001; Multimedia Appendix 1) and the desktop PC monitor (κ=0.663, *P*<.001; Multimedia Appendix 2). Interrater agreement was also favorable between the 2 vascular neurologists for the different devices (KS using the Join smartphone app and TK using the desktop PC monitor: κ=0.663, *P*<.001; KS using the desktop PC monitor and TK using the Join smartphone app: κ=0.723, *P*<.001; Multimedia Appendices 3 and 4).

Table 1. Patient characteristics (n=111).

Characteristic	Value
Age (years), median (IQR[a])	69 (58-78)
Male, n (%)	73 (66)
Past history, n (%)	
Hypertension	74 (68)
Hyperlipidemia	49 (44)
Diabetes mellitus	26 (23)
Atrial fibrillation	22 (20)
NIHSS[b] score on admission, median (IQR)	4 (2-7)
TOAST[c] classification, n (%)	
Large-artery atherosclerosis	8 (7)
Small-vessel occlusion	15 (14)
Cardioembolism	35 (32)
Other determined etiology	9 (8)
Undetermined	43 (40)
mRS[d] at 3 months, median (IQR)	1 (1-3)
Imaging	arge-artery atherosclerosis
DWI-ASPECTS[e], median (IQR)	9 (6-10)
Major arterial occlusion, n (%)	46 (41)
MRI[f] time from onset, median (IQR)	270 (185-335)

[a] IQR: interquartile range.

[b] NIHSS: National Institutes of Health Stroke Scale.

[c] TOAST: Trial of ORG 10172 in Acute Stroke Treatment.

[d] mRS: Modified Rankin Scale.

[e] DWI-ASPECTS: Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores.

[f] MRI: magnetic resonance imaging.

Figure 3. Scatter diagram of the DWI-ASPECTS results of vascular neurologists KS (A) and TK (B) between JOIN-ASPECTS and PC-ASPECTS for all patients. DWI-ASPECTS: Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores.

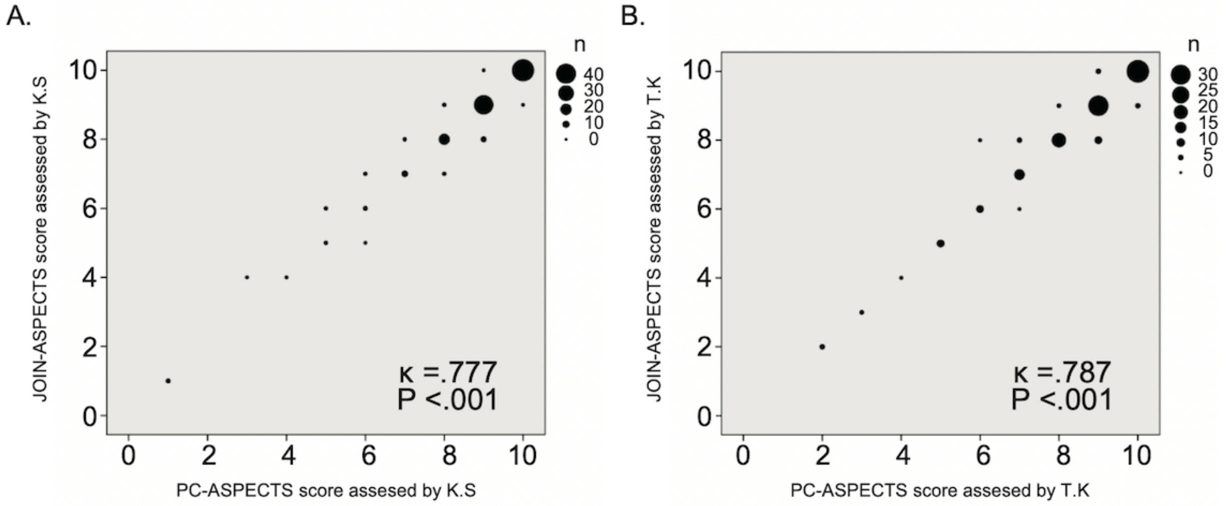


Figure 4. Scatter diagram of DWI-ASPECTS results of vascular neurologists KS (A) and TK (B) between JOIN-ASPECTS and PC-ASPECTS for patients without major artery occlusion. DWI-ASPECTS: Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores.

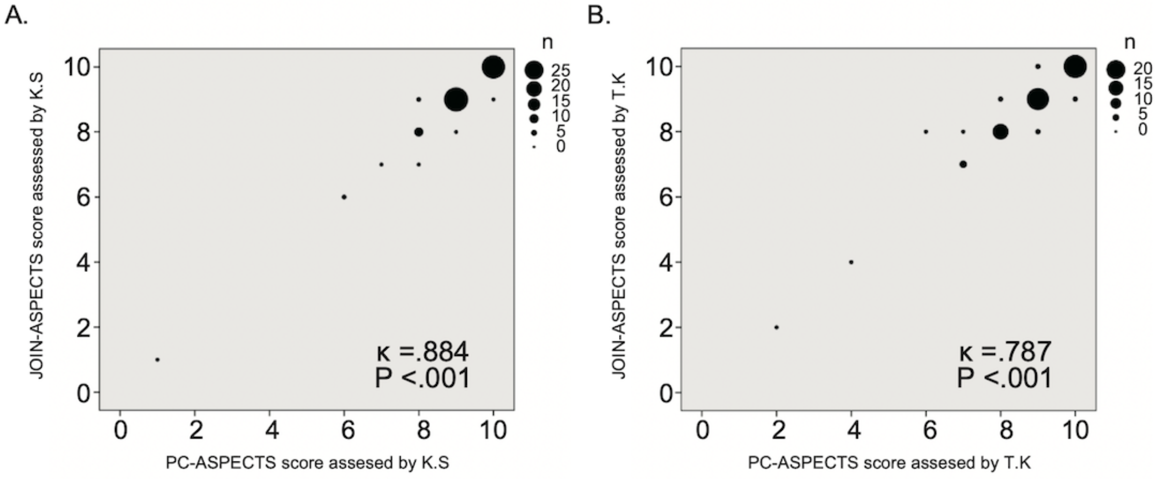
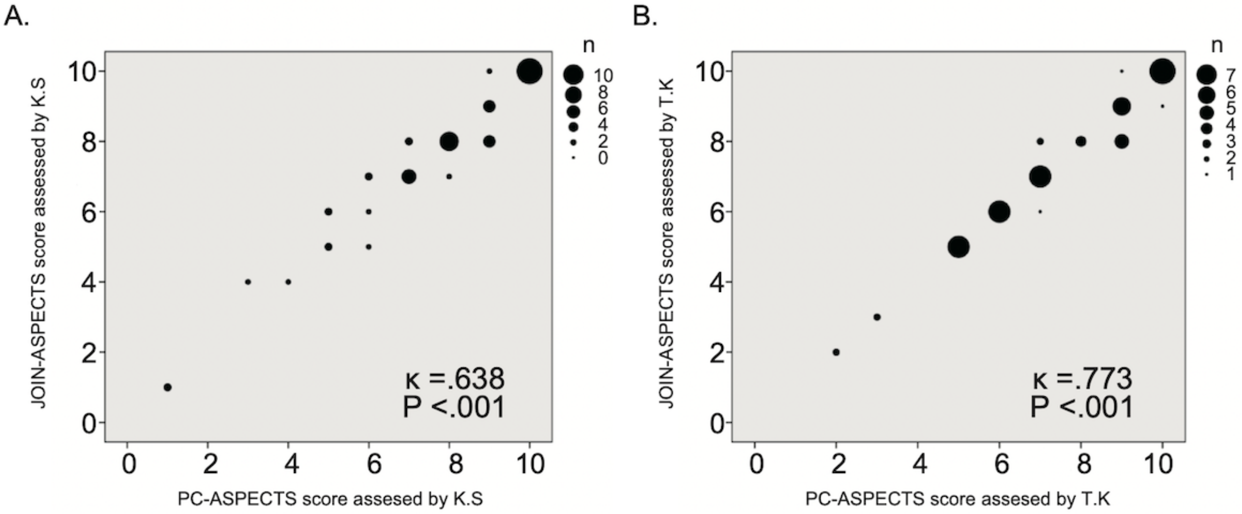


Figure 5. Scatter diagram of DWI-ASPECTS results of vascular neurologists KS (A) and TK (B) between JOIN-ASPECTS and PC-ASPECTS for 46 patients with major artery occlusion. DWI-ASPECTS: Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores.



Discussion

Principal Findings

There were 3 major findings in this study. First, there was a high degree of interdevice and interrater agreement in terms of the vascular neurologists' neuroimaging findings between the smartphone and the desktop PC monitor among patients with acute stroke. Second, DWI-ASPECTS was favorable for smartphone–desktop PC monitor, desktop PC monitor–desktop PC monitor, and smartphone–smartphone. Third, the elapsed interpretation time for DWI-ASPECTS using the smartphone was similar to that for the desktop PC monitor. The smartphone was comparable to the desktop PC monitor concerning DWI in acute ischemic stroke patients.

Our study presents the following 2 original points. First, we evaluated neuroimaging using 2 different devices, a smartphone, and a desktop PC monitor. Telemedicine has been proposed for assessment and treatment of acute stroke patients, but few studies have investigated the reliability of telemedicine for neuroimaging assessment.

Previous reports of hub-and-spoke type telemedicine networks have shown the validity and reliability of a telestroke neuroimaging system (desktop PC monitor–desktop PC monitor) in differentiating between ischemic stroke and hemorrhage stroke [9-11]. To the best of our knowledge, this study was the first to confirm the diagnostic accuracy of neuroimaging accessed using a smartphone app. In addition, we investigated the agreement of DWI-ASPECTS for ischemic lesions. Measurement of ischemic core volume is essential for hyperacute stroke therapy [2]. According to the findings of the DAWN [2] and WAKE-UP [4] trials, DWI imaging should feature more prominently in assessment of the suitability of hyperacute ischemic stroke patients for treatments such as thrombolysis. Therefore, DWI-ASPECTS is a useful and important scale in hyperacute stroke care. The score and the specific DWI cutoff value should be determined and shared before thrombolysis and mechanical thrombectomy.

Our results were in line with those of previous studies that reported almost perfect interrater agreement [12-14]. According to the findings of the DAWN and WAKE-UP trials, the therapeutic time for thrombolysis can be extended if a mismatch in visibility of a lesion is found between DWI and fluid-attenuated inversion recovery. Thrombolysis with perfusion imaging using a contrast agent, as in the DEFUSE 3 trial [3], is limited to some special comprehensive stroke centers.

DWI-ASPECTS was carefully evaluated and shared among the physicians who participated in this study. We expect that in the near future, assessment by DWI-ASPECTS and sharing of this information using the smartphone app will become commonplace in the management of hyperacute stroke patients.

It is crucial in telemedicine to have a high degree of interdevice agreement for DWI-ASPECTS between the smartphone and PC monitor scores. Sharing information such as neuroimaging among vascular neurologists, emergency department staff, and paramedical staff enables the medical team to deliver IV rtPA and mechanical thrombectomy in the shortest possible time.

In this study, there was excellent physician acceptance and a high level of satisfaction with the smartphone app system. One difference in this study was that our proposed neuroimaging telemedicine service can be conducted smoothly using various devices rather than being limited to the conventional PC–PC system. The security of personal patient details in the Join smartphone app enables the stroke team to safely and rapidly share information that is important for acute stroke therapy. Thus, we consider that the use of a medical smartphone app would change the manner of communication among the members of the stroke team.

Many physicians currently use smartphones daily as a part of their clinical examinations [15]. The main requirements of a telestroke consultation are fast and accurate neurological assessment by a stroke specialist; a review of brain imaging; and formulation of the diagnosis and treatment plan, including assessment of eligibility for standard thrombolytic therapy and endovascular devices. Telemedicine carries a large burden regarding investment in equipment, which typically includes two or more PCs, a web camera, network system, and software. The Join smartphone app could be used in place of all of these components, as a standalone tool or as an adjunct to existing telemedicine technology. The widespread use of smartphones, coupled with widely available health care apps, could enable the affordable expansion of telestroke networks. In addition to the Join smartphone app, numerous other videoconferencing and teleradiology apps are available in the smartphone app marketplace, many of which could facilitate telestroke consultation in the manner described here.

Our results revealed no significant difference in the time required for neuroimaging interpretation between the smartphone and desktop PC monitor systems. The main requirements of a telestroke consultation are rapid and accurate neurological assessment; review of neuroimaging; and formulation of the diagnosis and treatment plan, including IV rtPA and mechanical thrombectomy. We estimated that using a smartphone could take time for a neuroimaging evaluation, because this is a possibility due to a time lag resulting from connection issues. However, we found no delay in the time required for a neuroimaging review between the wireless connection by smartphone app and the wired connection by desktop PC monitor.

Smartphones are in common use by many physicians and already contribute to decisions made in medical treatment. We believe that the advantages of using a smartphone app are its portability and faster time to access to target neuroimages. Usually, stroke neurologists are not necessarily in front of a desktop or laptop computer. If an evaluation were requested, a stroke neurologist would need a certain amount of time to reach the computer, boot up the computer, log in to a network with a tight security system, and finally start browsing the images using a picture archiving and communication system. Using a smartphone app can solve the problem of the amount of time it takes a neurologist to evaluate neuroimaging by skipping these processes. Supplementation of the app with neuroimaging software has the potential to transform the smartphone into a complete tool for acute stroke evaluation. The Join smartphone app can play an important role in teleconsultation even when the stroke team comprises members who are located outside the hospital and who cannot access a desktop PC monitor.

There are several limitations in this study. First, the 2 vascular neurologists who evaluated the neuroimages have extensive clinical experience. It would be necessary to evaluate the interpretations of those physicians who have less experience. Second, we restricted our evaluation to neuroimaging of an anterior circulation stroke. It may be more difficult for vascular neurologists to detect a small infarction in the brain stem [16,17]. If applicable, axial and sagittal DWI should be routinely examined. Finally, our relatively high DWI-ASPECTS values compared with those of previous investigations [12-14] suggest the influence of high interdevice and interrater agreement in our results.

Conclusion

The Join smartphone app enabled stroke neurologists to estimate DWI-ASPECTS accurately and rapidly. We demonstrated the usefulness of the app in facilitating management of acute stroke using a system that offers sharing of images and discussion across different devices, in the manner of a social networking service.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Inter-rater agreement for DWI-ASPECTS ≥7 or DWI-ASPECTS <7 evaluated using the smartphone monitor.
[PDF File (Adobe PDF File), 10 KB - jmir_v22i6e15893_app1.pdf]

Multimedia Appendix 2
Table S2. Inter-rater agreement for DWI-ASPECTS ≥7 or DWI-ASPECTS <7 evaluated on a desktop PC monitor.
[PDF File (Adobe PDF File), 10 KB - jmir_v22i6e15893_app2.pdf]

Multimedia Appendix 3
Table S3. Inter-rater agreement for DWI-ASPECTS ≥7 or DWI-ASPECTS <7 for desktop PC monitor in VN1 vs smartphone monitor in VN2.
[PDF File (Adobe PDF File), 11 KB - jmir_v22i6e15893_app3.pdf]

Multimedia Appendix 4
Table S4. Inter-rater agreement for DWI-ASPECTS ≥7 or DWI-ASPECTS <7 for smartphone monitor in VN1 vs desktop PC monitor in VN2.
[PDF File (Adobe PDF File), 11 KB - jmir_v22i6e15893_app4.pdf]

Multimedia Appendix 5
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 334 KB - jmir_v22i6e15893_app5.pdf]

References

1. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. N Engl J Med 2008 Sep 25;359(13):1317-1329. [doi: 10.1056/NEJMoa0804656] [Medline: 18815396]
2. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, DAWN Trial Investigators. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. N Engl J Med 2018 Jan 04;378(1):11-21. [doi: 10.1056/NEJMoa1706442] [Medline: 29129157]
3. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, DEFUSE 3 Investigators. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. N Engl J Med 2018 Feb 22;378(8):708-718 [FREE Full text] [doi: 10.1056/NEJMoa1713973] [Medline: 29364767]
4. Thomalla G, Simonsen CZ, Boutitie F, Andersen G, Berthezene Y, Cheng B, WAKE-UP Investigators. MRI-Guided Thrombolysis for Stroke with Unknown Time of Onset. N Engl J Med 2018 Aug 16;379(7):611-622. [doi: 10.1056/NEJMoa1804355] [Medline: 29766770]
5. Campbell BCV, Mitchell PJ, Churilov L, Yassi N, Kleinig TJ, Dowling RJ, EXTEND-IA TNK Investigators. Tenecteplase versus Alteplase before Thrombectomy for Ischemic Stroke. N Engl J Med 2018 Apr 26;378(17):1573-1582. [doi: 10.1056/NEJMoa1716405] [Medline: 29694815]
6. Mitchell JR, Sharma P, Modi J, Simpson M, Thomas M, Hill MD, et al. A smartphone client-server teleradiology system for primary diagnosis of acute stroke. J Med Internet Res 2011 May 6;13(2):e31 [FREE Full text] [doi: 10.2196/jmir.1732] [Medline: 21550961]
7. Finlayson O, John V, Yeung R, Dowlatshahi D, Howard P, Zhang L, et al. Interobserver agreement of ASPECT score distribution for noncontrast CT, CT angiography, and CT perfusion in acute stroke. Stroke 2013 Jan;44(1):234-236. [doi: 10.1161/STROKEAHA.112.665208] [Medline: 23103490]
8. Nezu T, Koga M, Nakagawara J, Shiokawa Y, Yamagami H, Furui E, et al. Early ischemic change on CT versus diffusion-weighted imaging for patients with stroke receiving intravenous recombinant tissue-type plasminogen activator therapy: stroke acute management with urgent risk-factor assessment and improvement (SAMURAI) rt-PA registry. Stroke 2011 Aug;42(8):2196-2200. [doi: 10.1161/STROKEAHA.111.614404] [Medline: 21719764]
9. Spokoyny I, Raman R, Ernstrom K, Demaerschalk BM, Lyden PD, Hemmen TM, et al. Pooled assessment of computed tomography interpretation by vascular neurologists in the STRokE DOC telestroke network. J Stroke Cerebrovasc Dis 2014 Mar;23(3):511-515 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2013.04.023] [Medline: 23697761]
10. Demaerschalk BM, Bobrow BJ, Raman R, Ernstrom K, Hoxworth JM, Patel AC, Stroke Team Remote Evaluation Using a Digital Observation Camera (STRokE DOC) in Arizona—The Initial Mayo Clinic Experience (AZ TIME) Investigators. CT interpretation in a telestroke network: agreement among a spoke radiologist, hub vascular neurologist, and hub neuroradiologist. Stroke 2012 Nov;43(11):3095-3097 [FREE Full text] [doi: 10.1161/STROKEAHA.112.666255] [Medline: 22984007]
11. Puetz V, Bodechtel U, Gerber JC, Dzialowski I, Kunz A, Wolz M, et al. Reliability of brain CT evaluation by stroke neurologists in telemedicine. Neurology 2013 Jan 22;80(4):332-338. [doi: 10.1212/WNL.0b013e31827f07d0] [Medline: 23255831]

12. McTaggart RA, Jovin TG, Lansberg MG, Mlynash M, Jayaraman MV, Choudhri OA, DEFUSE 2 Investigators. Alberta stroke program early computed tomographic scoring performance in a series of patients undergoing computed tomography and MRI: reader agreement, modality agreement, and outcome prediction. Stroke 2015 Feb;46(2):407-412. [doi: 10.1161/STROKEAHA.114.006564] [Medline: 25538199]
13. Mitomi M, Kimura K, Aoki J, Iguchi Y. Comparison of CT and DWI Findings in Ischemic Stroke Patients within 3 Hours of Onset. Journal of Stroke and Cerebrovascular Diseases 2014 Jan;23(1):37-42. [doi: 10.1016/j.jstrokecerebrovasdis.2012.08.014]
14. Fiebach J, Schellinger P, Jansen O, Meyer M, Wilde P, Bender J, et al. CT and Diffusion-Weighted MR Imaging in Randomized Order. Stroke 2002 Sep;33(9):2206-2210. [doi: 10.1161/01.str.0000026864.20339.cb]
15. Demaerschalk BM, Vegunta S, Vargas BB, Wu Q, Channer DD, Hentz JG. Reliability of real-time video smartphone for assessing National Institutes of Health Stroke Scale scores in acute stroke patients. Stroke 2012 Dec;43(12):3271-3277 [FREE Full text] [doi: 10.1161/STROKEAHA.112.669150] [Medline: 23160878]
16. Loh P, Butcher KS, Parsons MW, MacGregor L, Desmond PM, Tress BM, et al. Apparent diffusion coefficient thresholds do not predict the response to acute stroke thrombolysis. Stroke 2005 Dec;36(12):2626-2631. [doi: 10.1161/01.STR.0000189688.95557.2b] [Medline: 16269645]
17. Bykowski JL, Latour LL, Warach S. More accurate identification of reversible ischemic injury in human stroke by cerebrospinal fluid suppressed diffusion-weighted imaging. Stroke 2004 May;35(5):1100-1106. [doi: 10.1161/01.STR.0000125867.86298.6a] [Medline: 15060314]

Edited by G Eysenbach; submitted 16.08.19; peer-reviewed by I Yuki, M Reynolds; comments to author 29.02.20; revised version received 05.03.20; accepted 23.03.20; published 09.06.20.

Please cite as:
Sakai K, Komatsu T, Iguchi Y, Takao H, Ishibashi T, Murayama Y
Reliability of Smartphone for Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores in Acute Ischemic Stroke Patients: Diagnostic Test Accuracy Study
J Med Internet Res 2020;22(6):e15893
URL: https://www.jmir.org/2020/6/e15893
doi:10.2196/15893
PMID:32515744

Original Paper

Menstrual Cycle Length and Patterns in a Global Cohort of Women Using a Mobile Phone App: Retrospective Cohort Study

Jessica A Grieger[1,2], BSc, PhD; Robert J Norman[1,3], MD, FRANZCOG

[1]Robinson Research Institute, University of Adelaide, Adelaide, Australia
[2]Adelaide Medical School, University of Adelaide, Adelaide, Australia
[3]Fertility SA, Adelaide, Australia

Corresponding Author:
Jessica A Grieger, BSc, PhD
Robinson Research Institute
University of Adelaide
North Terrace
Adelaide, 5000
Australia
Phone: 61 88137697
Email: jessica.grieger@adelaide.edu.au

Abstract

Background: There is increasing information characterizing menstrual cycle length in women, but less information is available on the potential differences across lifestyle variables.

Objective: This study aimed to describe differences in menstrual cycle length, variability, and menstrual phase across women of different ages and BMI among a global cohort of Flo app users. We have also reported on demographic and lifestyle characteristics across median cycle lengths.

Methods: The analysis was run based on the aggregated anonymized dataset from a menstrual cycle tracker and ovulation calendar that covers all phases of the reproductive cycle. Self-reported information is documented, including demographics, menstrual flow and cycle length, ovulation information, and reproductive health and diseases. Data from women aged ≥18 years and who had logged at least three cycles (ie, 2 completed cycles and 1 current cycle) in the Flo app were included (1,579,819 women).

Results: Of the 1.5 million users, approximately half (638,683/1,579,819, 40.42%) were aged between 18 and 24 years. Just over half of those reporting BMIs were in the normal range (18.5-24.9 kg/m^2; 202,420/356,598, 56.76%) and one-third were overweight or obese (>25 kg/m^2; 120,983/356,598, 33.93%). A total of 16.32% (257,889/1,579,819) of women had a 28-day median cycle length. There was a higher percentage of women aged ≥40 years who had a 27-day median cycle length than those aged between 18 and 24 years (22,294/120,612, 18.48% vs 60,870/637,601, 9.55%), but a lower percentage with a 29-day median cycle length (10,572/120,612, 8.77% vs 79,626/637,601, 12.49%). There were a higher number of cycles with short luteal phases in younger women, whereas women aged ≥40 years had a higher number of cycles with longer luteal phases. Median menstrual cycle length and the length of the follicular and luteal phases were not remarkably different with increasing BMI, except for the heaviest women at a BMI of ≥50 kg/m^2.

Conclusions: On a global scale, we have provided extensive evidence on the characteristics of women and their menstrual cycle length and patterns across different age and BMI groups. This information is necessary to support updates of current clinical guidelines around menstrual cycle length and patterns for clinical use in fertility programs.

(J Med Internet Res 2020;22(6):e17109) doi:10.2196/17109

KEYWORDS

body mass index; follicular phase; fertility; luteal phase; menstrual cycle; mobile application; internet; ovulation; mobile phone

Introduction

Background

For more than 30 years, the notion has been that the average woman is potentially fertile between days 10 and 17 of her 28-day menstrual cycle. However, this assumes that ovulation occurs exactly 14 days before the onset of the next menstrual period and that the fertile window extends before and after ovulation [1]. Data on aggregated cycles show mean cycle lengths of 28 to 29 days [2,3] and 29 to 30 days [4-7]. Other studies reporting on the number of women and their cycle length have demonstrated that 77% of women in the United States and Canada have a cycle length of 25 to 31 days [5]; 65% of rural Chinese women have a cycle length of 27 to 29 days [8]; and 84% of Australian women have a cycle length of 26 to 34 days [9]. Few studies also reported on the estimated day of ovulation, demonstrating an estimated day of ovulation at day 14 in only 3% to 10% of cycles [10] and at day 15 in 12% to 16% cycles [10].

Landmark longitudinal [7,11] and cross-sectional [5] studies on menstrual cycle variability in the late 1960s and 1970s demonstrated that the majority of women had a cycle length between 15 and 45 days [5,7,11], with mean cycle lengths decreasing with increasing age. Subsequent studies have reported similar findings with a shorter cycle length and less variability with increasing age [3,12,13]. Studies among Western populations have shown that longer menstrual cycles were associated with a higher BMI [14], increased parity [14], and recent use of oral contraceptives [15], whereas shorter menstrual cycles were associated with smoking [14,16], heavy caffeine intake [17], and alcohol consumption [18]. There is much less information on other lifestyle and behavioral factors; however, women reporting depression, higher perceived stress levels, and high levels of physical activity were associated with having irregular periods, anovulation, and heavier menstrual bleeding [19-22], highlighting the impact that a range of modifiable factors may have on future fertility. Additionally, although early studies with small sample sizes show measures of obesity and body composition associated with menstrual cycle irregularities [23-26], the specific impact of BMI has not been clearly determined. This is important given the global increase in BMI, of which the rise in obesity prevalence has been most prominent in women of reproductive age [27].

For the past 10 years, hundreds of smartphone apps have been developed for women to track their menstrual cycles, ovulation and fertile days, and health information [4,28-32]. A calendar-based method app used with 949 women indicated a mean 28-day cycle (range 17-35 days), with the most likely day of ovulation being day 16 [33], and in 45,360 women using the *Ovia Fertility* app, the mean cycle length was 30.4 (SD 4.6) days [34]. In over 600,000 cycles collected using the *Natural Cycles* app, the mean cycle length was 29.3 (SD 5.2) days [4], and in 225,596 cycles, the mean cycle length was 29.6 (SD 5.4) days [35]. Despite this surge in population data of menstrual cycle length, information on *individual women* is lacking rather than aggregates of cycle length, and it remains unclear how many women actually have an *average* 28-day cycle. Furthermore, the characteristics of women according to cycle length and lifestyle are limited.

Objectives

This study aimed to describe differences in menstrual cycle length, variability, and menstrual phase across women of different ages and BMIs. We also reported on demographic and lifestyle characteristics across the median cycle length. The study analyses will provide extensive worldwide evidence on the characteristics of menstrual cycle length and patterns among a global cohort of Flo app users. This information is necessary to support recommendations within current obstetric clinical guidelines around menstrual cycle length and patterns for clinical use in fertility programs.

Methods

Data Collection

The data were collected through the mobile period tracking and ovulation calendar called Flo, which was launched in October 2015. Users of the Flo app are found in over 200 countries, but the vast majority of them are located in Europe and North America. Flo’s privacy policy and terms of use permit the use of aggregated and anonymized data for research purposes. Data collection for this study started in December 2018 and extracted from 2 million users aged ≥18 years. Women were not included if they had logged a pregnancy, oral contraceptive reminders, and had recorded less than 3 menstrual cycles, which was the minimum number needed to obtain median cycle data. This resulted in a cohort of 1,579,819 women.

App Content

The participants' ages were collected through the mandatory sign-up questionnaire, and BMI was calculated for women who reported their weight (the latest input was used) and height. Additional information was collected from the user in 2 ways: (1) by tracking the symptoms that users log themselves and (2) by survey answer questions. For example, the user can log information on cycle length, period length and intensity, and days of the luteal phase. The survey information that users can report includes a range of questions regarding diet, lifestyle (general well-being, sleep, sexual activity, physical activity, stress, alcohol), menstrual cycle symptoms, and reproductive health information. Manually logged information reported in the current analyses included information on weight (in kilogram, logged manually), menstrual flow (light, medium, heavy), smoking status (regularly, sometimes, do not smoke), alcohol consumption (≥3 times a week, 1-2 times a week, <2-3 times a month, none), frequency of high stress (≥3 times a week, 1-2 times a week, <3 times a month, none), frequency of physical activity (≥3 times a week, 1-2 times a week, <3 times a month, none), relationship status (married or stable partner, single, no partner), months trying to conceive (<1 month, 1-3 months, 3-6 months, >6 months to 1 year, >1 year), number of children (0, 1, or 2, ≥3), and any reproductive disorders (yes, no, I don't know).

Menstrual Cycle Characteristics

Women manually logged information about their menstrual cycles (days of menstruation), including the intensity of menstrual flow (light, medium, heavy). The start and end dates of menstrual cycles were defined by the logged first day of menstruation. To estimate mean and median cycle lengths, we computed intrawomen inputs and then calculated the mean and median for population groups.

Menstruation flow calculation was completed for every cycle for each woman. If the length of a period was 1 or 2 days, it was calculated as low, but for a length of ≥3 days with more than half of the days logged with the intensity, it was calculated as light (L)=1, medium (M)=2, heavy (H)=3, none=no data. These numbers were arbitrary, as there is no measure of the volume of blood loss during menstruation. The average was computed for all bleeding days per cycle (eg, bleeding days: LMHHML=1+2+3+3+2+1=12 divided by 6=2=M). If the average was <1.5, it was recorded as light. If the average was ≥1.5 to <2.4, it was recorded as medium, and if the average was ≥2.4, it was recorded as heavy flow. In the case of ≥3 days logged and <50% days with reported intensities, it was considered as not tracked. Calculations were performed for all cycles that were logged by each woman.

Ovulation was recorded in women who reported a positive luteinizing hormone (LH) test, signifying the end of the follicular phase, and the next day estimated ovulation. In the case of multiple results logged, the last one was used. The length of the follicular phase was estimated from the first day of the menstrual cycle up to and excluding the day of ovulation. The length of the luteal phase was defined as the period from the day of ovulation up to and excluding the first day of menstruation.

The cycle length variation was calculated as the expected value of the mean square deviation, that is, var(x)=E[(x-mu)2], where E[.] is expectation and mu=E[x].

Results

Participant Characteristics

The data from over 1.5 million nonpregnant women who were not using oral contraceptives and who had logged at least 3 cycles (N=1,579,819) were included in the study (Table 1). The number of women with 2 completed and 1 current cycle was 123,245, with 3 to 10 cycles (n=819,933), 11 to 20 cycles (n=450,096), 21 to 30 cycles (n=154,536), and ≥31 cycles (n=32,009). Overall, the women were young, did not smoke, and rarely consumed alcohol but were mostly sedentary. That is, approximately half (638,683/1,579,819, 40.42%) of the users were aged 18 to 24 years. Only 22.57% (356,598/1,579,819) of the women reported their height and weight (to determine BMI); over half had a BMI in the normal range of 18.5 to 24.9 kg/m^2 (202,420/356,598, 56.76%) and one-third (120,983/356,598, 33.93%) were overweight or obese. About half of the total population reported their smoking, alcohol, and physical activity habits, of which 25.73% (192,118/746,396) smoked, 25.10% (186,378/742,615) consumed alcohol ≥1 time per week, and 45.36% (363,314/800,870) did not engage in physical activity.

Table 1. Description of the study population.

Characteristics	Values, n (%)
Age (years; n=1,579,819)	
18-24	638,683 (40.43)
25-29	383,179 (24.25)
30-34	275,039 (17.41)
35-39	162,156 (10.26)
40-55	120,762 (7.64)
BMI (kg/m^2; n=356,598)	
≤18.4	33,195 (9.31)
18.5-24.9	202,420 (56.76)
25.0-29.9	70,707 (19.83)
30.0-34.9	30,144 (8.45)
35.0-50.0	19,540 (5.48)
≥50+	592 (0.17)
Menstruation flow (n=197,579)	
Light	44,171 (22.36)
Medium	148,236 (75.03)
Heavy	5172 (2.62)
Smoking status (n=746,369)	
Regularly	104,612 (14.01)
Sometimes	87,506 (11.72)
Do not smoke	554,278 (74.26)
How often do you drink alcohol? (n=742,615)	
≥3 times a week	42,115 (5.67)
1-2 times a week	144,263 (19.42)
<2-3 times a month	310,597 (41.82)
None	245,640 (33.08)
Frequency of high stress (n=768,571)	
≥3 times a week	252,894 (32.90)
1-2 times a week	248,630 (32.35)
<3 times a month	218,490 (28.43)
None	48,557 (6.32)
Frequency of physical activity (n=800,870)	
≥3 times a week	140,287 (17.52)
1-2 times a week	176,124 (22.00)
<3 times a month	121,145 (15.13)
None	363,314 (45.36)
Relationship status (n=748,612)	
Married or stable partner	531,360 (70.98)
Single	73,194 (9.78)
No partner	144,058 (19.24)
Trying to conceive (n=137,973)	
<1 month	36,510 (26.46)

Characteristics	Values, n (%)
1-3 months	27,383 (19.85)
3-6 months	17,108 (12.40)
>6 months to 1 year	15,088 (10.94)
>1 year	41,884 (30.36)
Number of children (n=709,037)	
0	501,976 (70.80)
1 or 2	178,303 (25.15)
≥3	28,758 (4.06)
Any reproductive disorders (n=687,290)	
Yes	76,722 (11.16)
No	443,597 (64.54)
I don’t know	166,971 (24.29)

Menstrual Cycle Length and Characteristics

In all, 91.13% (1,439,613/1,579,819) of women had a median cycle length of 21 to 35 days, whereas 89.04% (1,406,643/1,579,819) had an average cycle length of 21 to 35 days. A total of 0.17% (2614/1,579,819) had a short cycle length (<21 days), and 8.60% (135,824/1,579,819) had a long cycle length (>35 days). The percentage of women with a median 27-day, 28-day, and 29-day cycle length was 12.05% (190,373/1,579,819), 16.32% (257,889/1,579,819), and 12.11% (191,351/1,579,819), respectively. Table 2 describes the demographic and lifestyle characteristics of the women, split by median cycle length. A higher percentage of women with short cycles reported having high stress ≥3 times/week (448/1117, 40.10%) and performing no physical activity (622/1158, 53.71%) compared with women with normal and long cycles. A higher percentage of women with short cycles had a lighter menstrual flow (240/476, 50.4%) compared with women with normal (37,469/172,025, 21.78%) and long cycles (6462/25,078, 25.77%). Consumption of alcohol or smoking status did not appear to influence menstrual cycle length (Table 2).

Figures 1 and 2 show the median cycle length distribution by age and BMI, respectively. In women aged 35 to 39 years and ≥40 years, more than 16% had a median cycle length of 27 days, compared with 9.6% of women aged 18 to 24 years. Comparatively, <8% of women aged 35 to 39 years and ≥40 years had a median 30-day cycle length compared with 10.8% of women aged 18 to 24 years. For BMI <50 kg/m^2, the highest percentage of women had a 28-day median cycle (15%-18%) and tended to track similarly across BMIs for shorter and longer day cycles (Figure 2). There were 592 women (0.17% of all women) with a BMI of ≥50 kg/m^2; the highest percentage had a median 29-day cycle (n=80, 13.6%). A higher percentage of women with a BMI of 35 to 50 kg/m^2 (11.6%) and >50 kg/m^2 (11.7%) had a median cycle of 36 days or more, compared with between 7.5% and 9.6% for the lower BMI categories.

Table 2. Demographic and lifestyle characteristics across median cycle length.

Characteristics	Short cycle (≤20 days), n (%)	Normal cycle (21-35 days), n (%)	Long cycle (≥36 days), n (%)
Age (years)[a]			
18-24	1255 (48.01)	556,541 (39.57)	80,887 (47.42)
25-29	634 (24.25)	334,817 (23.80)	47,728 (27.98)
30-34	330 (12.62)	248,280 (17.65)	26,429 (15.50)
35-39	170 (6.50)	151,867 (10.80)	10,119 (5.93)
40-55	225 (8.61)	115,138 (8.19)	5399 (3.17)
BMI (kg/m^2)[b]			
≤18.4	31 (8.1)	29,309 (9.21)	3855 (10.16)
18.5-24.9	182 (47.6)	182,511 (57.35)	19,727 (51.98)
25.0-29.9	85 (22.3)	63,183 (19.85)	7439 (19.60)
30.0-34.9	40 (10.5)	26,253 (8.25)	3851 (10.15)
35.0-50.0	42 (11.0)	16,515 (5.19)	2983 (7.86)
≥50.0	2 (0.5)	494 (0.16)	96 (0.25)
Menstrual flow[c]			
Light	240 (50.4)	37,469 (21.78)	6462 (25.77)
Medium	222 (46.6)	130,178 (75.67)	17,836 (71.12)
Heavy	14 (2.94)	4378 (2.54)	780 (3.11)
Smoking status[d]			
Smoke regularly	160 (14.6)	94,649 (14.18)	9803 (12.56)
Smoke sometimes	155 (14.1)	78,205 (11.72)	9146 (11.72)
Do not smoke	782 (71.3)	494,381 (74.09)	59,115 (75.73)
How often do you drink alcohol?[e]			
≥3 times a week	60 (5.5)	38,455 (5.79)	3600 (4.64)
1-2 times a week	213 (19.5)	130,651 (19.68)	13,399 (17.25)
<2-3 times a month	444 (40.7)	277,731 (41.84)	32,422 (41.75)
None	373 (34.2)	217,031 (32.69)	28,236 (36.36)
Frequency of high stress[f]			
≥3 times a week	448 (40.1)	224,951 (32.75)	27,495 (34.13)
1-2 times a week	359 (32.1)	221,845 (32.30)	26,426 (32.80)
<3 times a month	235 (21.0)	196,573 (28.62)	21,682 (26.92)
None	75 (6.7)	43,528 (6.34)	4954 (6.15)
Frequency of physical activity[g]			
≥3 times a week	191 (16.5)	127,088 (17.76)	13,008 (15.49)
1-2 times a week	190 (16.4)	158,190 (22.10)	17,744 (21.13)
<3 times a month	155 (13.4)	107,620 (15.04)	13,370 (15.92)
None	622 (53.7)	322,839 (45.11)	39,853 (47.46)
Trying to conceive[h]			
<1 month	63 (27.8)	32,361 (26.82)	4086 (23.93)
1-3 months	49 (1.6)	23,828 (19.75)	3506 (20.53)
3-6 months	35 (15.4)	14,902 (12.35)	2171 (12.71)

Characteristics	Short cycle (≤20 days), n (%)	Normal cycle (21-35 days), n (%)	Long cycle (≥36 days), n (%)
>6 months to 1 year	18 (7.9)	13,096 (10.85)	1974 (11.56)
>1 year	62 (27.3)	36,484 (30.23)	5338 (31.26)
Number of children[i]			
0	650 (63.22)	444,711 (70.14)	56,615 (76.55)
1 or 2	324 (31.52)	162,663 (25.65)	15,316 (20.71)
≥3	54 (5.25)	26,679 (4.21)	2025 (2.74)
Any reproductive disorders?[j]			
Yes	164 (16.7)	64,376 (10.47)	12,182 (17.01)
No	563 (57.2)	403,419 (65.63)	39,615 (55.32)
I don't know	257 (26.1)	146,896 (23.90)	19,818 (27.67)

[a] Short cycle (N=2614), normal cycle (N=1,406,643), and long cycle (N=170,562).

[b] Short cycle (N=382), normal cycle (N=318,265), and long cycle (N=37,951).

[c] Short cycle (N=476), normal cycle (N=172,025), and long cycle (N=25,078).

[d] Short cycle (N=1097), normal cycle (N=667,235), and long cycle (N=78,064).

[e] Short cycle (N=1090), normal cycle (N=663,868), and long cycle (N=77,657).

[f] Short cycle (N=1117), normal cycle (N=686,897), and long cycle (N=80,557).

[g] Short cycle (N=1158), normal cycle (N=715,737), and long cycle (N=83,975).

[h] Short cycle (N=227), normal cycle (N=120,671), and long cycle (N=17,075).

[i] Short cycle (N=1028), normal cycle (N=634,053), and long cycle (N=73,956).

[j] Short cycle (N=984), normal cycle (N=614,691), and long cycle (N=71,615).

Figure 1. Median cycle length distribution by age groups.

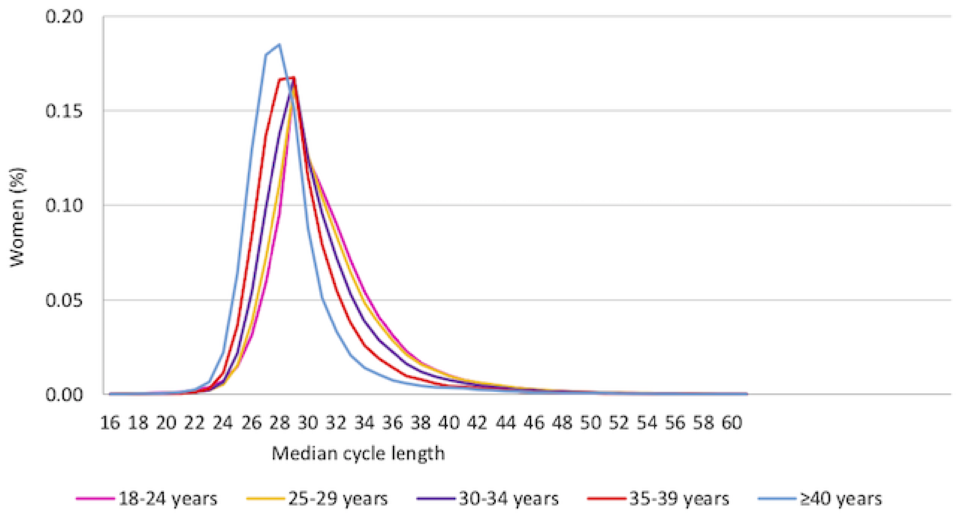
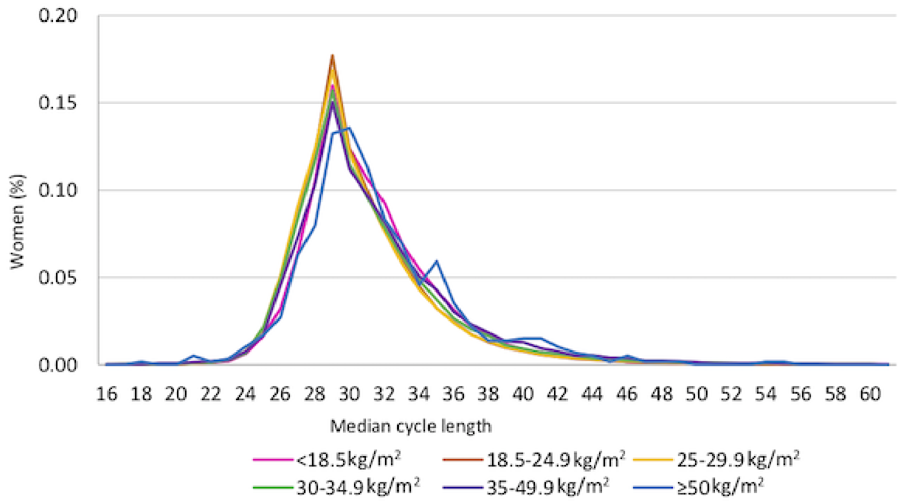


Figure 2. Median cycle length distribution by BMI categories.



Cycle Length Variability

Figure 3 illustrates the variation in cycle length for all women and for each age group. Among all women, 25.37% (275,715/1,086,923) had a cycle length variation of 0 to 1.5 days and 69% (753,831/1,086,923) of women had a <6-day variation. Variation in cycle length of 1.5 to 4.5 days was more common for women aged ≥35 years than for women aged <29 years, and this was particularly prominent at 1.5 days variation. That is, in women aged 35 to 39 years, and ≥40 years, a 1.5-day variation in cycle length was found in 25.17% (27,754/110,252) and 26.64% (14,255/53,502) of women, respectively, compared with 18.08% (63,314/350,021) and 20.12% (45,683/227,109) of women aged between 18 and 24 years and 25 and 29 years, respectively. Normal cycle variation between 3 and 4.5 days was also higher in women aged between 35 and 39 years (16.11%; 17,761/110,252) and ≥40 years (16.85%; 9015/53,502) compared with women aged between 18 and 24 years (13.38%; 46,817/350,021) and 25 and 29 years (14.17%; 32,185/227,109). Comparatively, a variation of ≥6 days was more prominent in the younger women and tracked similarly with increasing age. Across BMI groups, women with a higher BMI tended to have less variation in their cycle. This was particularly evident with a cycle variation of ≥4.5 days, where there were fewer women with a higher BMI and more women with a low BMI (<18 kg/m^2; Figure 4).

Figure 3. Variation in cycle length across age group.

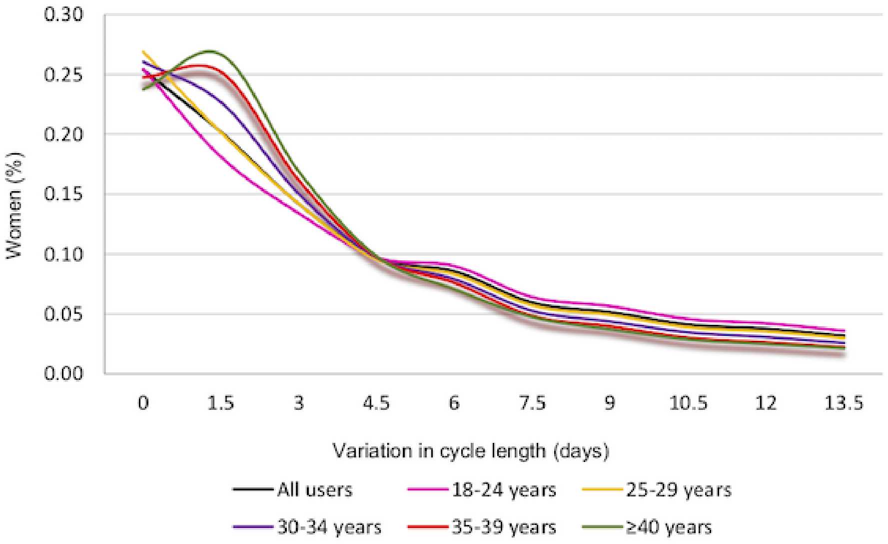
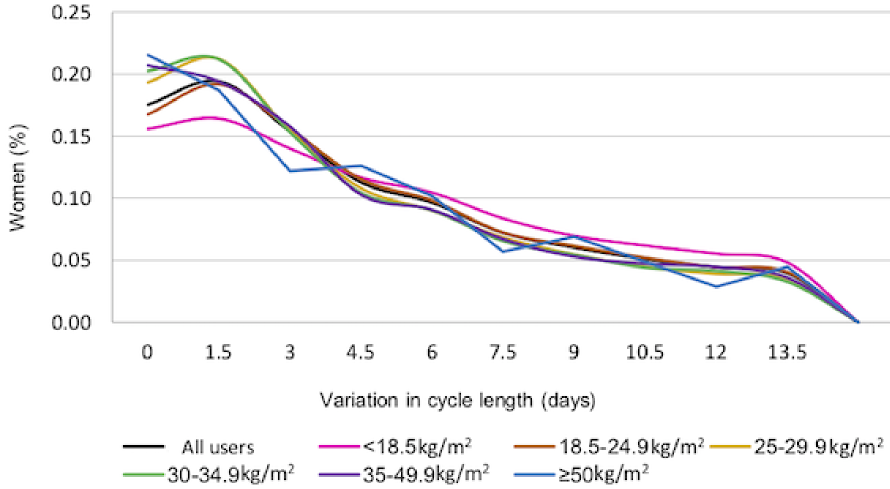


Figure 4. Variation in cycle length across BMI categories.



Ovulation and Phases of the Menstrual Cycle

There were 18,761 cycles with positive LH ovulation tests, signifying the end of the follicular phase, and the following day to be the day of estimated ovulation. The highest percentage of cycles was at an estimated ovulation day 14 (13.08%, 2509 cycles; Table 3), indicating a 13-day follicular phase (Multimedia Appendix 1). There was a similar percentage of cycles with estimated ovulation at day 13 (2248 cycles, 12.13%) and day 15 (2267 cycles, 11.96%). The number of cycles with estimated ovulation continued to reduce outside these days. Across age groups (Multimedia Appendix 2), the highest percentage of cycles was at an estimated ovulation day 15 for women aged between 18 and 24 years (187/1687 cycles, 11.08%), day 14 for women aged between 25 and 39 years (2108 cycles, 12%-15%), but at day 12 for women aged 40 years (237/1594 cycles, 14.9%). Across BMI categories (Multimedia Appendix 3), more frequent positive ovulation tests occurred at day 14, indicating a 13-day follicular length, except for a BMI of 35 to 49.9 kg/m^2, which was highest at day 13 (80 cycles).

The length of the luteal phase was highest at day 15, with 16.96% (3696/21,788) of the cycles, followed by a 14-day luteal phase, with 16.17% (3523/21,788) of the cycles (Table 4). The number of cycles reduced to around 12% at a 13-day and 16-day luteal phase length and dropped to 6% to 7% at a 12-day and 17-day luteal phase length. The luteal phase length across age and BMI are reported in Multimedia Appendices 4 and 5, respectively. For all age groups, the number of cycles was highest for a 15-day luteal phase length, followed by a 14-day luteal phase length. There was a higher percentage of cycles with shorter (5 to 10 day) luteal phases among the youngest compared with the higher age groups, but for a luteal phase of 13 to 16 days, more cycles were present in the age group of 25 years. Comparatively, there was a higher percentage of cycles from women aged ≥40 years that had longer luteal phases (≥15 days; Multimedia Appendix 4). A luteal phase length of 15 days was most common across all BMI categories except for the 30.0 to 34.9 kg/m^2 category, which was at a length of 14 days. Overall, the percentage of cycles tended to track similarly across BMI categories; however, there tended to be a higher percentage of cycles with a luteal phase of 11 to 13 days from overweight (25.0 to 29.9 kg/m^2) and obese (30.0 to 50.0 kg/m^2) women compared with the other BMI categories (Multimedia Appendix 5).

Table 3. Estimated ovulation day based on positive luteinizing hormone test (N=18,761 cycles).

Estimated day of ovulation	Cycles, n (%)
11	1668 (8.89)
12	1877 (10.00)
13	2248 (11.98)
14	2509 (13.37)
15	2267 (12.08)
16	1905 (10.15)
17	1518 (8.09)
18	1154 (6.15)
19	822 (4.38)
20	636 (3.39)
21	458 (2.44)
22	321 (1.71)
23	304 (1.62)
24	266 (1.41)
25	189 (1.00)
26	184 (0.98)
27	149 (0.79)
28	160 (0.85)
29	126 (0.67)

Table 4. Estimated luteal phase length based on positive luteinizing hormone test (N=18,761).

Length of luteal phase (days)	Cycles, n (%)
5	142 (0.75)
6	172 (0.91)
7	206 (1.09)
8	227 (1.20)
9	344 (1.83)
10	543 (2.89)
11	946 (5.04)
12	1565 (8.34)
13	2645 (14.09)
14	3523 (18.77)
15	3696 (19.70)
16	2765 (14.73)
17	1819 (9.69)
18	1380 (7.35)
19	1006 (5.36)
20	809 (4.31)

Discussion

Principal Findings

We have reported recent and comprehensive data characterizing over 1.5 million women and their menstrual cycle patterns. Approximately two-thirds of women using the app were aged <30 years and one-third were overweight or obese based on their BMI. The majority of women (1,439,613/1,579,819, 1.13%) had a usual median cycle length of 21 to 35 days, of whom 16.32% (257,889/1,579,819) had a median cycle length of 28 days. Shorter cycles with less variation were more common for women ≥40 years. Among women who reported short cycles, a higher percentage reported higher stress levels, no physical activity, and a lighter menstrual flow in comparison with women with normal or long cycles. There were only subtle differences in length and variability across BMIs; however, women with the lowest BMI tended to have the highest variation in cycle length and more women with a BMI >35 kg/m^2 had a median cycle of 36 days and longer.

Strengths and Limitations

The reporting of menstrual cycle length and patterns was estimated from mobile phone data collected from women using the app. Although this app has a global reach of 100 million women using the app, and 30 million active users monthly, not all women reported on their BMI or other lifestyle characteristics. This means these findings cannot be generalized to all women using the app. Recently, a review of pregnancy apps and their use in culturally and linguistically diverse women found that engagement with apps may be lower in areas of higher social disadvantage, cultural and language barriers, and health literacy [36]. This study did not ask about reasons for using the app (except if it was to become pregnant); thus, it is unclear whether women who provided more demographic and fertility details have a different level of knowledge or interest in reproductive health. We did not have data to assess cultural disparities across menstrual cycle characteristics, which would be valuable in future research. Information on cycle intensity was self-reported based on subjective feelings of *light*, *medium*, and *heavy* flow; objective measures of menstrual flow would improve the accuracy of cycle intensity. We used the data on women to assess menstrual cycle characteristics, but also used the number of cycles to assess ovulation. Understanding the characteristics of women that contribute to anovulatory cycles would provide important information on other potential issues with fertility. Among women who have reproductive health issues, including polycystic ovary syndrome or endometriosis, the ability to become pregnant may already be reduced. An app estimating ovulation or suggesting the most fertile period for those women may be limited in its effectiveness.

The strengths of this study include the very large cohort of women providing fertility information, along with data on a high number of ovulatory cycles. The Flo app uses artificial intelligence–based algorithms, and when additional data, such as basal body temperature and ovulation data, are entered, the predictive ability improves. Comparatively, calendar-based apps may only use simulated 28-day cycles to estimate ovulation, or predictive methods may be used; for example, the standard days method, where users avoid unprotected intercourse during cycle days 8 through 19 [37] or the rhythm method, which predicts fertile days using a formula and is based on data from the menstrual records of the past 6 cycles [6]. They do not provide information on how the predictions are calculated or the accuracy of predictions.

Comparison With Prior Work

Imperative to our study is that we assessed *individual* women, and we report on the number and percentage of women and median menstrual cycle length. Our results are useful in the clinic setting and provide information on the actual occurrence of cycle data in women but also differences in cycle patterns across age groups, BMI, and some lifestyle characteristics. Importantly, we show that a cycle length of 28 days is not common for many women, which is usually the reference length for a typical woman. Comparatively, previous analyses among women using fertility apps or cohort studies have reported that the mean or median cycle length for all women ranged between 28 and 30 days [8,11,33,34], and studies reporting on actual cycles reported means between 27.7 and 29.6 days [2,4,5,7,35]. These results cannot be extrapolated to an individual woman, and clinically, it is not helpful, as we have clearly demonstrated that the percentage of women with an average cycle length of 28 days is uncommon. Our results provoke a potential paradigm shift, changing our thinking from the current concept of mean cycle length to frequencies of cycle length in individual women.

Our results are consistent with previous studies reporting on shorter cycle lengths with increasing age [3-5,11], but also less variation with older age [5,13]. Additionally, there is a dependence between age and length of the luteal phase. Only a small number of cycles were assessed in our luteal phase data, but our results are partly consistent with the study by Lenton et al [12] showing a higher incidence of shorter cycles in the youngest women (18-24 years) but also the oldest women aged between 45 and 50 years. A recent study by Bull et al [4] using a fertility app did not show any difference in the luteal phase across age groups. Such differences are likely because of the assessment of this phase, and when using an app, one that includes thermal changes, but also objective markers of ovulation are more likely to provide accurate information.

Epidemiological evidence has demonstrated the adverse effects of obesity on female reproduction, including anovulation and menstrual cycle irregularities [24,25,38], impaired pregnancy success using assisted reproductive technology, and infertility [39]. Furthermore, mechanistic studies have confirmed alterations in reproductive hormones associated with obesity [40]. Small studies demonstrated body composition measures associated with menstrual cycle irregularities [23-26], and an Australian study of 726 women was the first to show that obese women were twice as likely to have greater cycle variation (>15 days) compared with women with a normal weight [9]. In a recent app study using >600,000 cycles, there were no differences in the percentage of cycles across BMI categories, except morbidly obese women had higher cycle length variation by 0.4 days and longer follicular phase length by 0.9 days, than women with a normal BMI [4]. The current analysis revealed that median menstrual cycle length and length of follicular and luteal phases were not remarkably different with increasing BMI, except for the heaviest women at a BMI of ≥50 kg/m^2. It is interesting to note that of the over 1.5 million women in the study, only 22.57% (356,598/1,579,819) reported height and weight to calculate BMI. Although this might demonstrate some information bias such that our findings may systematically underestimate the relationship between BMI and menstrual cycle, the proportion of women who were overweight or obese (33.93%) in this study is similar to the current prevalence rates in the general and reproductive population of women [27,41]. Thus, the effect of BMI on menstrual cycle length may be mediated by other factors that we did not assess in our descriptive study. Given the inconsistent findings and lack of clear association between BMI and menstrual cycle characteristics in large cohorts of women (rather than data on cycles), further investigation into this area is needed.

This is the largest analysis to report on physical activity, stress, and menstrual flow across short, normal, and long menstrual cycle lengths. Interestingly, shorter cycles were associated with a higher frequency of high stress, no physical activity, and lighter menstrual flow. There is limited information on stress and physical activity, but research to date suggests that professional and high-frequency training impairs ovarian activity, which may manifest as luteal phase defects, irregular menstruation, or amenorrhea [42-44]. In nontrained women aged 18 to 44 years, total recreational physical activity and vigorous recreational activity were positively associated with cycle length, but not cycle variability [45]. Regarding sedentary behavior, among 2613 nulliparous Danish women, those who were sedentary (ie, <5 metabolic equivalents [METs] per week of physical activity), had a higher prevalence of irregular cycles than women who engaged in moderate levels of activity (20 to 39 METs per week; probability ratio 1.54; 95% CI 1.16-2.04) [46]. Small cross-sectional studies have shown that higher stress is a significant predictor of irregular menstrual cycles [47] and dysmenorrhea [48]. Although these findings support the benefits of moderate physical activity and minimizing stress, further investigation into the impact of lifestyle factors on menstrual cycle characteristics is necessary.

The current findings have clinical implications for women who are trying to conceive but also for women with different lifestyle characteristics. First, current estimations for conceiving are based on the premise of a 28-day cycle with ovulation on day 14. We demonstrated that only 16% of women had a median 28-day cycle. Our results invoke a fundamental change in the assessment of cycle lengths of individual women and that the population mean cycle length may no longer be an applicable form of its measurement. Second, few menstrual cycle differences were seen in women with different BMIs. Finally, lifestyle factors such as physical activity and stress may relate to cycle length and therefore the potential day of ovulation. Prediction models, including menstrual cycle data as well as age and lifestyle factors, could be developed to predict the likelihood of pregnancy success. The use of such models would be helpful in the clinic setting to optimize conception, particularly in those who do not have regular cycles.

Conclusions

On a global scale, we have provided extensive evidence on the characteristics of women and their menstrual cycle length and patterns. We demonstrate that the typical average 28-day cycle length is not common for a high percentage of women, and only 13.08% of cycles had an estimated ovulation on day 14. Age appeared to be a more important factor than BMI in terms of menstrual cycle length and variability. Our data provide important information on the necessity for an individualized approach to support reproductive health and fertility, and modifiable factors such as physical activity and stress should additionally be considered when planning a pregnancy. Future work could extend these findings by addressing cultural and ethnic diversities in relation to menstrual cycle patterns.

Acknowledgments

The authors gratefully acknowledge the members of the Flo team, Anna Targonskaya, Natalia Kukharchyk, Anna Klepchukova, and Andrei Zhabinsk, who generated the results and allowed the authors to make use of the data. Without them, this paper would not have been possible. The authors appreciate their guidance and assistance with the results and the overall paper.

Authors' Contributions

JG and RN contributed substantially to the conception and design of the study. JG wrote the first draft of the manuscript. All authors read the final draft of the manuscript and revised it critically for its intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Estimated length of the follicular phase.
[PDF File (Adobe PDF File), 94 KB - jmir_v22i6e17109_app1.pdf]

Multimedia Appendix 2
Estimated ovulation day based on positive luteinizing hormone test across age groups (N=18,761 cycles).

[PDF File (Adobe PDF File), 51 KB - jmir_v22i6e17109_app2.pdf]

Multimedia Appendix 3

Estimated ovulation day based on positive luteinizing hormone test across BMI group (N=18,235 cycles).

[PDF File (Adobe PDF File), 58 KB - jmir_v22i6e17109_app3.pdf]

Multimedia Appendix 4

Number of cycles and the luteal length, according to age.

[PDF File (Adobe PDF File), 103 KB - jmir_v22i6e17109_app4.pdf]

Multimedia Appendix 5

Number of cycles and the luteal length, according to BMI.

[PDF File (Adobe PDF File), 84 KB - jmir_v22i6e17109_app5.pdf]

References

1. Beckmann CL, Hebert W, Laube D, Smith R, Barzansky B. Obstetrics and Gynecology. Third Edition. Baltimore, USA: Williams and Wilkins; 1998.
2. World Health Organization. Temporal relationships between indices of the fertile period. Fertil Steril 1983 May;39(5):647-655. [doi: 10.1016/s0015-0282(16)47060-3] [Medline: 6840308]
3. Fehring RJ, Schneider M, Raviele K. Variability in the phases of the menstrual cycle. J Obstet Gynecol Neonatal Nurs 2006;35(3):376-384. [doi: 10.1111/j.1552-6909.2006.00051.x] [Medline: 16700687]
4. Bull JR, Rowland SP, Scherwitzl EB, Scherwitzl R, Danielsson KG, Harper J. Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. NPJ Digit Med 2019;2(1):83 [FREE Full text] [doi: 10.1038/s41746-019-0152-7] [Medline: 31482137]
5. Chiazze L, Brayer FT, Macisco JJ, Parker MP, Duffy BJ. The length and variability of the human menstrual cycle. J Am Med Assoc 1968 Feb 5;203(6):377-380. [Medline: 5694118]
6. Creinin MD, Keverline S, Meyn LA. How regular is regular? An analysis of menstrual cycle regularity. Contraception 2004 Oct;70(4):289-292. [doi: 10.1016/j.contraception.2004.04.012] [Medline: 15451332]
7. Vollman RF. The menstrual cycle. Major Probl Obstet Gynecol 1977;7:1-193. [Medline: 836520]
8. Zhang Q, Wang Y, Zhang Y, Zhang H, Yang Y, He Y, et al. The influence of age at menarche, menstrual cycle length and bleeding duration on time to pregnancy: a large prospective cohort study among rural Chinese women. BJOG 2017 Oct;124(11):1654-1662 [FREE Full text] [doi: 10.1111/1471-0528.14469] [Medline: 28128508]
9. Wei S, Schmidt M, Dwyer T, Norman R, Venn A. Obesity and menstrual irregularity: associations with SHBG, testosterone, and insulin. Obesity (Silver Spring) 2009 May;17(5):1070-1076 [FREE Full text] [doi: 10.1038/oby.2008.641] [Medline: 19180069]
10. Baird DD, McConnaughey DR, Weinberg CR, Musey PL, Collins DC, Kesner JS, et al. Application of a method for estimating day of ovulation using urinary estrogen and progesterone metabolites. Epidemiology 1995 Sep;6(5):547-550. [doi: 10.1097/00001648-199509000-00015] [Medline: 8562634]
11. Treloar AE, Boynton RE, Behn BG, Brown BW. Variation of the human menstrual cycle through reproductive life. Int J Fertil 1967;12(1 Pt 2):77-126. [Medline: 5419031]
12. Lenton EA, Landgren BM, Sexton L, Harper R. Normal variation in the length of the follicular phase of the menstrual cycle: effect of chronological age. Br J Obstet Gynaecol 1984 Jul;91(7):681-684. [doi: 10.1111/j.1471-0528.1984.tb04830.x] [Medline: 6743609]
13. Münster K, Schmidt L, Helm P. Length and variation in the menstrual cycle-a cross-sectional study from a Danish county. Br J Obstet Gynaecol 1992 May;99(5):422-429. [doi: 10.1111/j.1471-0528.1992.tb13762.x] [Medline: 1622917]
14. Kato I, Toniolo P, Koenig KL, Shore RE, Zeleniuch-Jacquotte A, Akhmedkhanov A, et al. Epidemiologic correlates with menstrual cycle length in middle aged women. Eur J Epidemiol 1999 Oct;15(9):809-814. [doi: 10.1023/a:1007669430686] [Medline: 10608360]
15. Jukic AM, Weinberg CR, Baird DD, Wilcox AJ. Lifestyle and reproductive factors associated with follicular phase length. J Womens Health (Larchmt) 2007 Nov;16(9):1340-1347 [FREE Full text] [doi: 10.1089/jwh.2007.0354] [Medline: 18001191]
16. Windham GC, Elkin EP, Swan SH, Waller KO, Fenster L. Cigarette smoking and effects on menstrual function. Obstet Gynecol 1999 Jan;93(1):59-65. [doi: 10.1016/s0029-7844(98)00317-2] [Medline: 9916957]
17. Fenster L, Quale C, Waller K, Windham GC, Elkin EP, Benowitz N, et al. Caffeine consumption and menstrual function. Am J Epidemiol 1999 Mar 15;149(6):550-557. [doi: 10.1093/oxfordjournals.aje.a009851] [Medline: 10084244]
18. Liu Y, Gold EB, Lasley BL, Johnson WO. Factors affecting menstrual cycle characteristics. Am J Epidemiol 2004 Jul 15;160(2):131-140. [doi: 10.1093/aje/kwh188] [Medline: 15234934]

19. Ansong E, Arhin SK, Cai Y, Xu X, Wu X. Menstrual characteristics, disorders and associated risk factors among female international students in Zhejiang Province, China: a cross-sectional survey. BMC Womens Health 2019 Feb 18;19(1):35 [FREE Full text] [doi: 10.1186/s12905-019-0730-5] [Medline: 30777053]
20. Nillni Y, Wesselink A, Hatch E, Mikkelsen E, Gradus JL, Rothman KJ, et al. Mental health, psychotropic medication use, and menstrual cycle characteristics. Clin Epidemiol 2018;10:1073-1082 [FREE Full text] [doi: 10.2147/CLEP.S152131] [Medline: 30214312]
21. Schliep KC, Mumford SL, Vladutiu CJ, Ahrens KA, Perkins NJ, Sjaarda LA, et al. Perceived stress, reproductive hormones, and ovulatory function: a prospective cohort study. Epidemiology 2015 Mar;26(2):177-184 [FREE Full text] [doi: 10.1097/EDE.0000000000000238] [Medline: 25643098]
22. Vannuccini S, Fondelli F, Clemenza S, Galanti G, Petraglia F. Dysmenorrhea and heavy menstrual bleeding in elite female athletes: quality of life and perceived stress. Reprod Sci 2020 Mar;27(3):888-894. [doi: 10.1007/s43032-019-00092-7] [Medline: 32046446]
23. Castillo-Martínez L, López-Alvarenga JC, Villa AR, González-Barranco J. Menstrual cycle length disorders in 18- to 40-y-old obese women. Nutrition 2003 Apr;19(4):317-320. [doi: 10.1016/s0899-9007(02)00998-x] [Medline: 12679164]
24. Harlow S, Matanoski G. The association between weight, physical activity, and stress and variation in the length of the menstrual cycle. Am J Epidemiol 1991 Jan;133(1):38-49. [doi: 10.1093/oxfordjournals.aje.a115800] [Medline: 1983897]
25. Hartz AJ, Barboriak PN, Wong A, Katayama KP, Rimm AA. The association of obesity with infertility and related menstural abnormalities in women. Int J Obes 1979;3(1):57-73. [Medline: 528119]
26. Rogers J, Mitchell GW. The relation of obesity to menstrual disturbances. N Engl J Med 1952 Jul 10;247(2):53-55. [doi: 10.1056/nejm195207102470204] [Medline: 14941282]
27. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the global burden of disease study 2013. Lancet 2014 Aug 30;384(9945):766-781 [FREE Full text] [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
28. Freis A, Freundl-Schütt T, Wallwiener L, Baur S, Strowitzki T, Freundl G, et al. Plausibility of menstrual cycle apps claiming to support conception. Front Public Health 2018;6:98 [FREE Full text] [doi: 10.3389/fpubh.2018.00098] [Medline: 29666788]
29. Jennings V, Haile LT, Simmons RG, Spieler J, Shattuck D. Perfect- and typical-use effectiveness of the Dot fertility app over 13 cycles: results from a prospective contraceptive effectiveness trial. Eur J Contracept Reprod Health Care 2019 Apr;24(2):148-153. [doi: 10.1080/13625187.2019.1581164] [Medline: 30880509]
30. Lupton D. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. Cult Health Sex 2015;17(4):440-453. [doi: 10.1080/13691058.2014.920528] [Medline: 24917459]
31. Starling MS, Kandel Z, Haile L, Simmons RG. User profile and preferences in fertility apps for preventing pregnancy: an exploratory pilot study. Mhealth 2018;4:21 [FREE Full text] [doi: 10.21037/mhealth.2018.06.02] [Medline: 30050917]
32. Symul L, Wac K, Hillard P, Salathé M. Assessment of menstrual health status and evolution through mobile apps for fertility awareness. NPJ Digit Med 2019;2:64 [FREE Full text] [doi: 10.1038/s41746-019-0139-4] [Medline: 31341953]
33. Johnson S, Marriott L, Zinaman M. Can apps and calendar methods predict ovulation with accuracy? Curr Med Res Opin 2018 Sep;34(9):1587-1594. [doi: 10.1080/03007995.2018.1475348] [Medline: 29749274]
34. Bradley D, Landau E, Jesani N, Mowry B, Chui K, Baron A, et al. Time to conception and the menstrual cycle: an observational study of fertility app users who conceived. Hum Fertil (Camb) 2019 May 16:1-9 epub ahead of print. [doi: 10.1080/14647273.2019.1613680] [Medline: 31094573]
35. Faust L, Bradley D, Landau E, Noddin K, Farland LV, Baron A, et al. Findings from a mobile application-based cohort are consistent with established knowledge of the menstrual cycle, fertile window, and conception. Fertil Steril 2019 Sep;112(3):450-7.e3. [doi: 10.1016/j.fertnstert.2019.05.008] [Medline: 31272722]
36. Hughson JP, Daly JO, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
37. Arévalo M, Jennings V, Sinai I. Efficacy of a new method of family planning: the standard days method. Contraception 2002 May;65(5):333-338. [doi: 10.1016/s0010-7824(02)00288-3] [Medline: 12057784]
38. Zain MM, Norman RJ. Impact of obesity on female fertility and fertility treatment. Womens Health (Lond) 2008 Mar;4(2):183-194. [doi: 10.2217/17455057.4.2.183] [Medline: 19072520]
39. Norman RJ, Clark AM. Obesity and reproductive disorders: a review. Reprod Fertil Dev 1998;10(1):55-63. [doi: 10.1071/r98010] [Medline: 9727593]
40. Leeners B, Geary N, Tobler P, Asarian L. Ovarian hormones and obesity. Hum Reprod Update 2017 May 1;23(3):300-321 [FREE Full text] [doi: 10.1093/humupd/dmw045] [Medline: 28333235]
41. Inoue Y, Qin B, Poti J, Sokol R, Gordon-Larsen P. Epidemiology of obesity in adults: latest trends. Curr Obes Rep 2018 Dec;7(4):276-288 [FREE Full text] [doi: 10.1007/s13679-018-0317-8] [Medline: 30155850]
42. Constantini NW, Dubnov G, Lebrun CM. The menstrual cycle and sport performance. Clin Sports Med 2005 Apr;24(2):e51-82, xiii. [doi: 10.1016/j.csm.2005.01.003] [Medline: 15892917]

43. Czajkowska M, Plinta R, Rutkowska M, Brzęk A, Skrzypulec-Plinta V, Drosdzol-Cop A. Menstrual cycle disorders in professional female rhythmic gymnasts. Int J Environ Res Public Health 2019 Apr 25;16(8):1470 [FREE Full text] [doi: 10.3390/ijerph16081470] [Medline: 31027248]
44. Meczekalski B, Katulski K, Czyzyk A, Podfigurna-Stopa A, Maciejewska-Jeske M. Functional hypothalamic amenorrhea and its influence on women's health. J Endocrinol Invest 2014 Nov;37(11):1049-1056 [FREE Full text] [doi: 10.1007/s40618-014-0169-3] [Medline: 25201001]
45. Sternfeld B, Jacobs M, Quesenberry CJ, Gold E, Sowers M. Physical activity and menstrual cycle characteristics in two prospective cohorts. Am J Epidemiol 2002 Sep 1;156(5):402-409. [doi: 10.1093/aje/kwf060] [Medline: 12196309]
46. Hahn K, Wise L, Riis AH, Mikkelsen EM, Rothman KJ, Banholzer K, et al. Correlates of menstrual cycle characteristics among nulliparous Danish women. Clin Epidemiol 2013;5:311-319 [FREE Full text] [doi: 10.2147/CLEP.S46712] [Medline: 23983490]
47. Yamamoto K, Okazaki A, Sakamoto Y, Funatsu M. The relationship between premenstrual symptoms, menstrual pain, irregular menstrual cycles, and psychosocial stress among Japanese college students. J Physiol Anthropol 2009;28(3):129-136 [FREE Full text] [doi: 10.2114/jpa2.28.129] [Medline: 19483374]
48. Kollipaka R, Arounassalame B, Lakshminarayanan S. Does psychosocial stress influence menstrual abnormalities in medical students? J Obstet Gynaecol 2013 Jul;33(5):489-493. [doi: 10.3109/01443615.2013.782272] [Medline: 23815203]

Abbreviations

H: heavy
L: light
LH: luteinizing hormone
M: medium
MET: metabolic equivalent

Edited by G Eysenbach; submitted 12.03.20; peer-reviewed by N Heynsbergh, A Teles; comments to author 08.04.20; revised version received 16.04.20; accepted 20.04.20; published 24.06.20.

Please cite as:
Grieger JA, Norman RJ
Menstrual Cycle Length and Patterns in a Global Cohort of Women Using a Mobile Phone App: Retrospective Cohort Study
J Med Internet Res 2020;22(6):e17109
URL: http://www.jmir.org/2020/6/e17109/
doi:10.2196/17109
PMID:32442161

Original Paper

Stratification of Individual Symptoms of Contact Lens–Associated Dry Eye Using the iPhone App DryEyeRhythm: Crowdsourced Cross-Sectional Study

Takenori Inomata[1,2,3], MBA, MD, PhD; Masahiro Nakamura[4], MD, PhD; Masao Iwagami[5], MPH, MSc, MD, PhD; Akie Midorikawa-Inomata[3], RN, MPH, PhD; Jaemyoung Sung[1,6], MS; Keiichi Fujimoto[7], MD; Yuichi Okumura[2,7], MD; Atsuko Eguchi[3]; Nanami Iwata[7], MS; Maria Miura[7], MD; Kenta Fujio[7], MD; Ken Nagino[3]; Satoshi Hori[8], MD, PhD; Kazuo Tsubota[9], MBA, MD, PhD; Reza Dana[10], MD, MPH, MSc; Akira Murakami[1], MD, PhD

[1]Department of Ophthalmology, Juntendo University Faculty of Medicine, Tokyo, Japan
[2]Department of Strategic Department of Operating Room Management and Improvement, Juntendo University Faculty of Medicine, Tokyo, Japan
[3]Department of Hospital Administration, Juntendo University Graduate School of Medicine, Tokyo, Japan
[4]Department of Bioengineering, Precision Health, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan
[5]Department of Health Services Research, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan
[6]Morsani College of Medicine, University of South Florida, Tampa, FL, United States
[7]Department of Ophthalmology, Juntendo University Graduate School of Medicine, Tokyo, Japan
[8]Department of Electric Medical Intelligence Management, Juntendo University Faculty of Medicine, Tokyo, Japan
[9]Department of Ophthalmology, Keio University School of Medicine, Tokyo, Japan
[10]Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Takenori Inomata, MBA, MD, PhD
Department of Ophthalmology
Juntendo University Faculty of Medicine
3-1-3 Hongo
Bunkyo-ku
Tokyo, 1130033
Japan
Phone: 81 3 3813 3111
Fax: 81 3 5689 0394
Email: tinoma@juntendo.ac.jp

Abstract

Background: Discontinuation of contact lens use is mainly caused by contact lens–associated dry eye. It is crucial to delineate contact lens–associated dry eye's multifaceted nature to tailor treatment to each patient's individual needs for future personalized medicine.

Objective: This paper aims to quantify and stratify individual subjective symptoms of contact lens–associated dry eye and clarify its risk factors for future personalized medicine using the smartphone app DryEyeRhythm (Juntendo University).

Methods: This cross-sectional study included iPhone (Apple Inc) users in Japan who downloaded DryEyeRhythm. DryEyeRhythm was used to collect medical big data related to contact lens–associated dry eye between November 2016 and January 2018. The main outcome measure was the incidence of contact lens–associated dry eye. Univariate and multivariate adjusted odds ratios of risk factors for contact lens–associated dry eye were determined by logistic regression analyses. The t-distributed Stochastic Neighbor Embedding algorithm was used to depict the stratification of subjective symptoms of contact lens–associated dry eye.

Results: The records of 4454 individuals (median age 27.9 years, SD 12.6), including 2972 female participants (66.73%), who completed all surveys were included in this study. Among the included participants, 1844 (41.40%) were using contact lenses, and among those who used contact lenses, 1447 (78.47%) had contact lens–associated dry eye. Multivariate adjusted odds ratios of risk factors for contact lens–associated dry eye were as follows: younger age, 0.98 (95% CI 0.96-0.99); female sex, 1.53 (95% CI 1.05-2.24); hay fever, 1.38 (95% CI 1.10-1.74); mental illness other than depression or schizophrenia, 2.51 (95% CI 1.13-5.57); past diagnosis of dry eye, 2.21 (95% CI 1.63-2.99); extended screen exposure time >8 hours, 1.61 (95% CI 1.13-2.28); and smoking, 2.07 (95% CI 1.49-2.88). The t-distributed Stochastic Neighbor Embedding analysis visualized and stratified 14 groups based on the subjective symptoms of contact lens–associated dry eye.

Conclusions: This study identified and stratified individuals with contact lens–associated dry eye and its risk factors. Data on subjective symptoms of contact lens–associated dry eye could be used for prospective prevention of contact lens–associated dry eye progression.

(J Med Internet Res 2020;22(6):e18996) doi:10.2196/18996

KEYWORDS

contact lens-associated dry eye; mobile health; ResearchKit; smartphone app; DryEyeRhythm; subjective symptoms; risk factors; dry eye; stratification; mobile phone

Introduction

Contact lens (CL) wear is an established and efficient method for improving vision quality by correcting refractive errors. More than 140 million estimated CL users exist worldwide [1]. However, despite available CL products on the market, studies show that 12% to 58% of CL users discontinue CL use due to CL discomfort (CLD) [1-6]. Additionally, many continue using CLs while feeling CLD.

Dry eye disease (DED) is characterized by a tear film disorder, potentially causing ocular surface damage and ocular discomfort [7,8]. DED is becoming more prevalent due to aging society and increased digital device usage [9-13]. Accumulating studies indicate that dryness is one of the main reasons for CLD [2,3,8]; 68% to 79% of CL users have reported feeling dryness [14-16], and CL use especially is a risk factor for severe DED [17,18]. CL-associated dry eye (CLADE) may be caused by changes of tear film stability on the CL surface, decrease in the tear exchange rate, decrease in reflex secretion due to perception decline, oxygen deprivation, lens deposits, and adverse reactions to CL solutions [5]. Suggested CLADE risk factors, including environmental and host factors, and lifestyle habits are related [1,19-21]; thus, a comprehensive, multidisciplinary mass customization is needed in individual CLADE treatments. Accordingly, it is crucial to understand CLADE's multifaceted nature by monitoring various symptoms and visualizing patient lifestyle practices to improve the CL users' quality of vision through personalized treatment [22,23]. Notably, a smartphone app could effectively monitor subjective symptoms and lifestyle habits, check changes in each factor's contribution, and visualize each individual’s lifestyle [22]. The collected medical data could help lay the foundation for understanding how individual factors contribute to the aggravation of CLADE.

To reveal and simplify how multiple factors intertwine to affect CLADE's progression, we conducted this large-scale crowdsourced study using the DryEyeRhythm app (Juntendo University) to quantify and stratify the symptoms of CLADE and collect evidence for prospective prevention of CLADE progression.

Methods

Study Enrollment and Participants

The DryEyeRhythm app's development and the study's enrollment process have been previously described [18,24,25]. Briefly, DryEyeRhythm was developed using Apple Inc’s open-source framework, ResearchKit. DryEyeRhythm was released on Apple’s App Store in Japan on November 2, 2016, and in the United States in April 2018 [18]. Prospective participants can download the app using their own App Store credentials. This large-scale, crowdsourced, prospective, cross-sectional observational study was conducted between November 2, 2016, and January 12, 2018. All users provided electronic informed consent for participation following explanation of the study's nature and possible consequences. Duplicate users, foreign participants (outside of Japan), and users who did not complete all surveys were excluded. This study was approved by the Independent Ethics Committee of Juntendo University Faculty of Medicine (approval number 19-226) and adhered to the tenets of the Declaration of Helsinki. The methodology and results of this survey are reported according to the checklist for reporting results of internet e-surveys [26].

User Data Collection

Using DryEyeRhythm, data on participant demographics, medical history, lifestyle habits, daily subjective symptoms, Ocular Surface Disease Index (OSDI) scores (Allergan), and Zung Self-Rating Depression Scale (SDS) scores were collected [27-29]. Multimedia Appendices 1 and 2 show specific questions and parameters used for data collection. Data that were recorded included basic demographic characteristics, including age, sex, height, body weight, race, and geographic location; medical history of hypertension, diabetes, systemic disease, mental illness, past diagnosis of DED, hay fever, and ophthalmic surgeries; and lifestyle habits, including daily coffee intake, CL information (such as the type of CL, period of CL use, and daily duration of CL use), eye drop use, screen exposure time, periodic exercise, sleep duration, smoking, and hydration. Daily subjective symptoms included eye itching, asthenopia, headache, mental fatigue, stiffness, and stress.

Classification of CLADE

Participants were divided into the following 2 groups: non-CLADE and CLADE. Those who reported current use of CL and had an OSDI total score <13 were included in the non-CLADE group. Those who reported current use of CL and had an OSDI total score ≥13 were included in the CLADE group.

The OSDI questionnaire is a 12-item questionnaire used to assess DED severity based on ocular symptoms, impact on visual functioning, and environmental triggers [27,29]. The overall OSDI total score was determined based on a 100-point scale correlated with the severity of symptoms [30]. We previously demonstrated that the Japanese version of OSDI with DryEyeRhythm had good validity compared with that with the paper-based questionnaire [18,29,31].

Depressive symptoms were evaluated using SDS [28]. The SDS is an internationally used 20-item self-administered depression scale and has been validated in Japan. Each item is rated on a 4-point Likert scale, with a total score ranging from 20 to 80. An SDS score of ≥40 is possibly suggestive of depression [32,33].

Statistical Analysis

Continuous variables (not normally distributed based on Shapiro-Wilk tests) are presented as medians (with interquartile ranges), and categorical variables are presented as percentages. We conducted Mann-Whitney U tests for continuous variables not normally distributed and chi-square tests for categorical variables. A comparison between negative, current, and past CL use groups was performed by one-way analysis of variance using a Bonferroni post hoc test. The odds ratio of each risk factor for CLADE was determined by multivariate adjusted logistic regression analysis, which included factors significantly associated with CLADE, as indicated by the univariable logistic regression analyses with a threshold 2-tailed, unpaired *P* value of .05. Pearson’s rank correlation coefficients were calculated to determine the correlation between each subjective symptom and CLADE. A heatmap was then made using the heatmap function of the seaborn module (version 0.9.0; Python 3). A t-distributed Stochastic Neighbor Embedding (t-SNE) was performed with a scikit-learn Python package (version 0.21.3; Python 3) [34]. *P* values were considered statistically significant at *P*<.05, *P*<.01, or *P*<.001. All data were analyzed with Stata (version 15; Stata Corp) software.

Results

Application Downloads and Clinical Study Enrollment

As seen in Figure 1, DryEyeRhythm was cumulatively downloaded 18,991 times between November 2, 2016, and January 12, 2018. As Figure 2 shows, a total of 21,394 records were identified in our crowd database; 11,485 and 5455 records were excluded from the study because of duplicate user data and incomplete survey responses, respectively. Finally, 4454 out of 9909 participants (44.95%) completed the survey and were included in the final analysis. Multimedia Appendix 3 shows the sensitivity analysis between included and excluded participants.

Figure 1. Trends in the number of downloads.

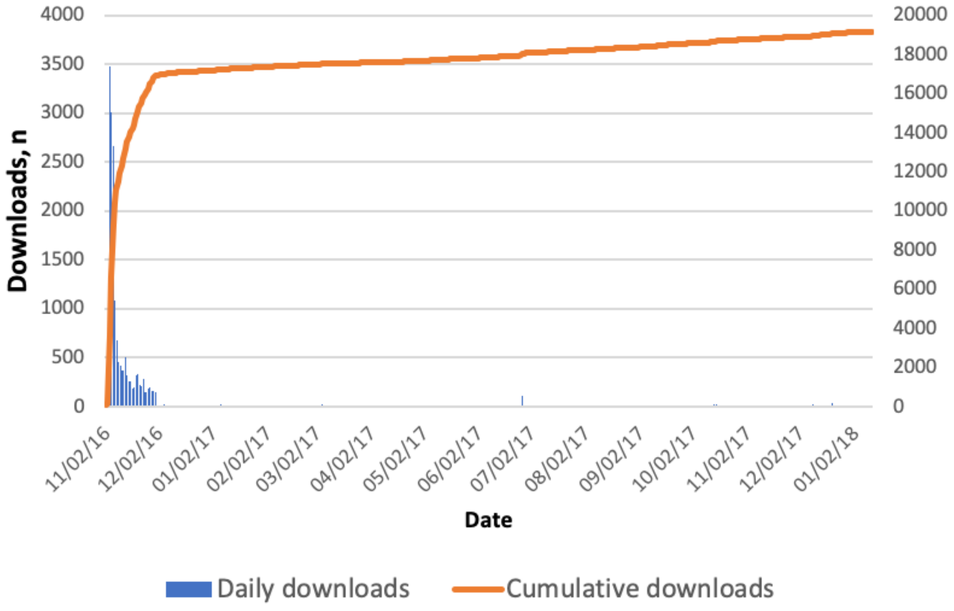
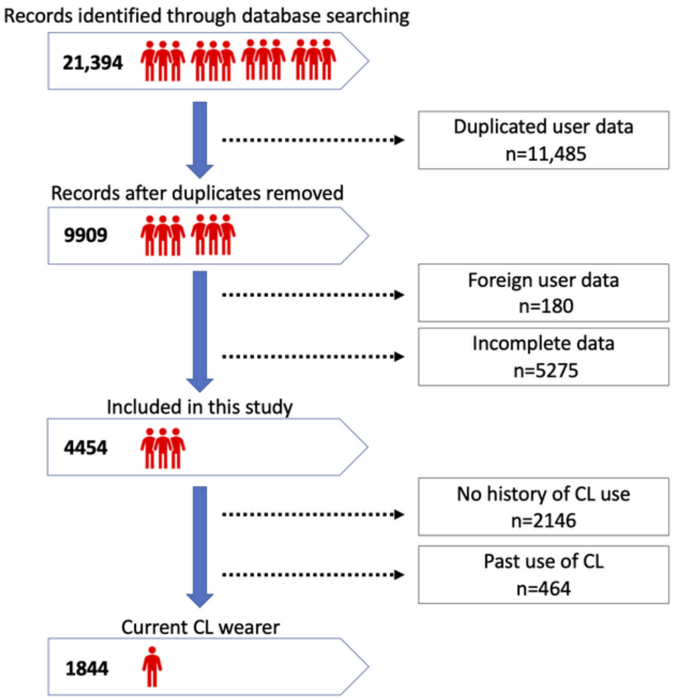


Figure 2. Flow chart of the enrollment process used in this study. CL: contact lens.



Participant Characteristics

Multimedia Appendix 4 summarizes participants' demographics, medical history, and lifestyle habits according to CL usage. The included participants' median age was 23 years (IQR 18-35), and women accounted for 66.73% (2972/4454) of the participants. Most participants were in the age group of 18 to 34 years (2362/4454 participants, 53.03%), followed by the age group of 35 to 64 years (1133/4454 participants, 25.44%). CL users were divided into the following subgroups: negative (2146/4454 participants, 48.18%), past use (464/4454 participants, 10.42%), and current use (1844/4454 participants, 41.40%). Most of the current CL users were younger than 65 years.

Regarding the medical history survey, the majority of the participants (2747/4454, 61.67%) did not present medicated hypertension, diabetes, systemic diseases, or mental illness history. Hay fever was present in about half of the participants (2249/4454, 50.49%). Previous DED diagnoses were reported in 26.08% of past (178/464) and current (481/1844) CL users, while 17.99% (386/2146 out of participants) of negative contact lens users had previous DED diagnoses. The history of laser-assisted in situ keratomileusis surgery was significantly increased in past CL users (*P*<.001).

Regarding lifestyle habits, eye drop use was reported in 23.49% (109/464) of past CL users and 26.03% (480/1844) of current CL users, while 13.42% (288/2146) of negative contact lens users did not use eye drops. Median screen exposure time was 6 hours (IQR 4-10). About two-thirds of participants (2919/4454, 65.54%) showed periodic exercise habits, with an average exercise time of 1 hour (IQR 0-3) per week. The median sleeping time was 7 hours (6-8.5) per night; however, more than half of participants (2355/4454, 52.87%) slept fewer than 6 hours per night. About a quarter (1058/4454, 23.75%) of participants smoked.

Characteristics of Non-CLADE and CLADE Groups

Multimedia Appendix 5 shows the characteristics of non-CLADE and CLADE groups. Most participants in both non-CLADE and CLADE groups were in the 18 to 34 years age group. There were significantly more female participants in the CLADE group (*P*<.001). Height and body weight were higher in the non-CLADE group. The median CL term was 6 hours (IQR 3-10) per day in both the non-CLADE and CLADE groups, and approximately half of the participants used CL for 12 to 18 hours per day. More than 90% (1725/1844, 93.55%) of the participants in both non-CLADE and CLADE groups used soft contact lenses (SCL). The majority of the SCL types were the daily disposable type (823/1844 participants, 44.63%), followed by the biweekly (every 2 weeks) type (741/1844 participants, 40.18%). Regarding the medical history survey, the prevalence of hay fever (*P*=.002), past diagnosis of DED (*P*<.001), and mental illnesses other than depression or schizophrenia (*P*=.003) were significantly higher in the CLADE group than in the non-CLADE group. Regarding lifestyle habits, eye drop use (*P*<.001), screen exposure time (*P*<.001), and smoking habits (*P*=.001) were significantly higher in the CLADE group than in the non-CLADE group.

Subjective Symptoms

Table 1 shows daily subjective symptoms and OSDI and SDS data. Eye itching was significantly higher in the CLADE group than in the non-CLADE group (*P*<.001). Other subjective symptoms, including asthenopia, headache, mental fatigue, and stiffness, were higher in the CLADE group compared with the non-CLADE group. The median OSDI total scores were 8.3 (IQR 6.3-10.4) in the non-CLADE group and 30 (IQR 20.8-43.8) in the CLADE group. Subscale scores of OSDI were higher in the CLADE group than in the non-CLADE group. The SDS score was also higher in the CLADE group, and 80.10% (1159/1447) of participants in this group experienced depressive symptoms (SDS score ≥40).

Table 1. Subjective symptoms of contact lens–associated dry eye.

Daily subjective symptoms	Non-CLADE[a] (n=397)	CLADE (n=1447)	*P* value
Eye itching (0-10), median (IQR)	0 (0-2)	2 (0-5)	<.001
Asthenopia (yes), n (%)	147 (37.0)	899 (62.1)	<.001
Headache (0-10), median (IQR)	0 (0-2)	1 (0-4)	<.001
Mental fatigue (yes), n (%)	57 (14.4)	446 (30.8)	<.001
Stiffness and pain of body axis muscle (yes), n (%)	151 (38.0)	816 (56.4)	<.001
Stress (1-10), median (IQR)	4 (2-6)	5 (3-7)	<.001
J-OSDI[b] total score (1-100), median (IQR)	8.3 (6.3-10.4)	30 (20.8-43.8)	<.001
Ocular symptoms (0-100), median (IQR)	15 (5-20)	35 (25-45)	<.001
(1) Eyes that are sensitive to light (0-4), median (IQR)	1 (0-1)	1 (1-3)	<.001
(2) Eyes that feel gritty (0-4), median (IQR)	1 (0-1)	1 (1-2)	<.001
(3)Painful or sore eyes (0-4), median (IQR)	0 (0-1)	1 (1-2)	<.001
(4) Blurred vision (0-4), median (IQR)	0 (0-1)	1 (1-2)	<.001
(5) Poor vision (0-4), median (IQR)	0 (0-1)	1 (1-2)	<.001
Vision-related function (0-100), median (IQR)	0 (0)	12.5 (6.3-25)	<.001
(6) Reading (0-4), median (IQR)	0 (0)	1 (0-1)	<.001
(7) Driving at night (0-4), median (IQR)	0 (0)	0 (0-1)	<.001
(8) Working with a computer or bank machine (0-4), median (IQR)	0 (0)	0 (0-1)	<.001
(9) Watching TV (0-4), median (IQR)	0 (0)	1 (0-1)	<.001
Environmental triggers (0-100), median (IQR)	8.3 (8.3-16.7)	41.7 (25-66.7)	<.001
(10) Windy conditions (0-4), median (IQR)	0 (0)	1 (1-3)	<.001
(11) Places or areas with low humidity (0-4), median (IQR)	0 (0-1)	2 (1-3)	<.001
(12) Areas that are air conditioned (0-4), median (IQR)	0 (0-1)	2 (1-3)	<.001
SDS[c] score (20-80), median (IQR)	42 (36-49)	48 (41-55)	<.001
SDS score >40, n (%)	234 (58.9)	1159 (80.1)	<.001

[a] CLADE: contact lens–associated dry eye.

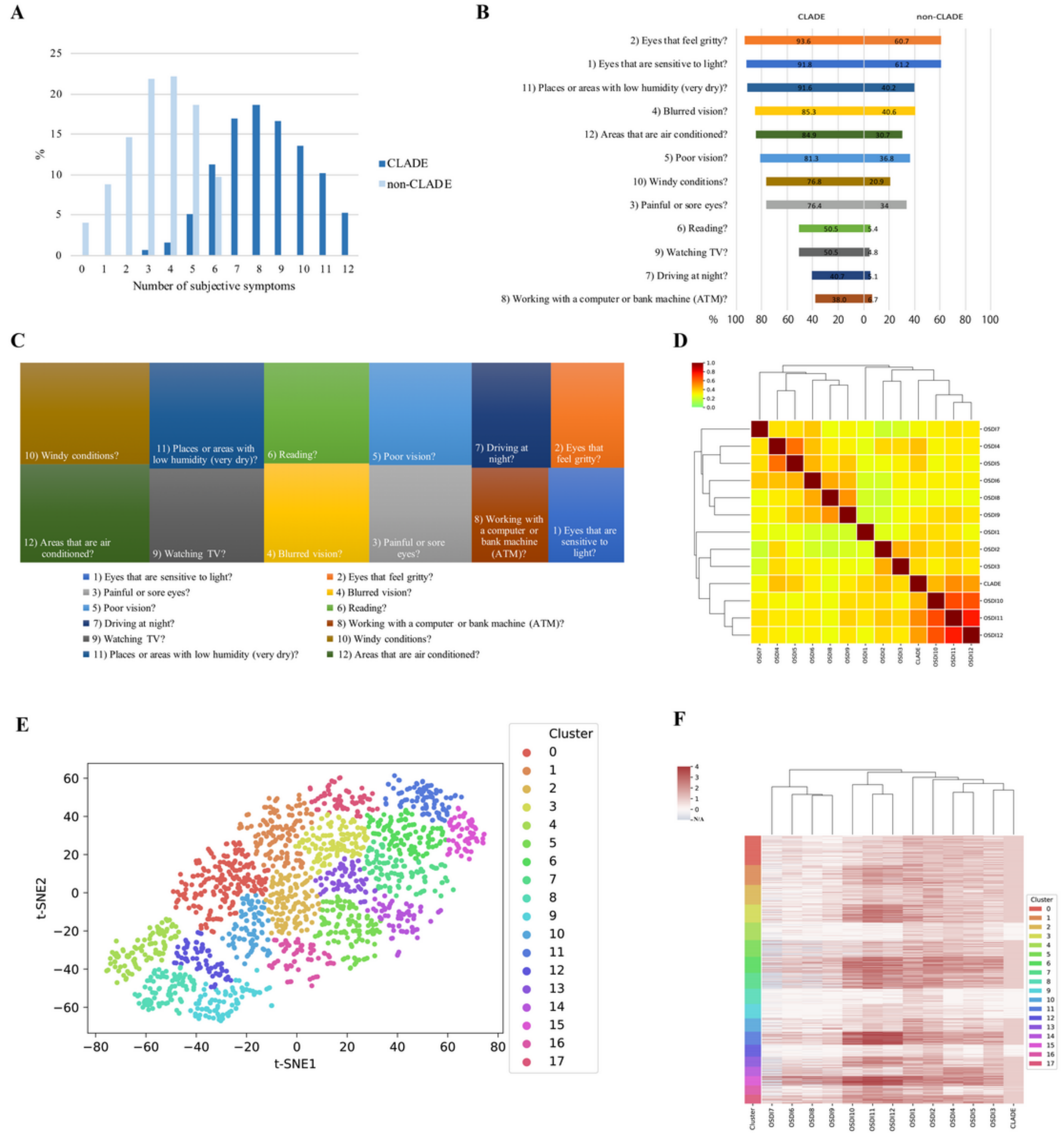
[b] J-OSDI: Japanese Ocular Surface Disease Index.

[c] SDS: Zung Self-Rating Depression Scale.

Stratification of Subjective Symptoms of CLADE

Figure 3 shows various representations of the stratification of the subjective symptoms of CLADE. The number of subjective symptoms for both CLADE and non-CLADE groups is based on OSDI questionnaires, and the number of subjective symptoms was higher in the CLADE group. The rate of positive signs in each OSDI item between the CLADE and non-CLADE groups shows that over 90% of CLADE participants felt gritty eyes, were sensitive to light, and felt uncomfortable in low-humidity places. The treemap consists of the differences between the CLADE and non-CLADE groups in the frequency of subjective symptoms of DED triggered by environmental factors (OSDI questions 10-12). The environmental factors are characterized as subjective symptoms of CLADE. The heatmap that visualizes the correlation between each subjective symptom and CLADE demonstrates that the subjective symptoms triggered by environmental factors (OSDI questions 10-12) were highly positively correlated with CLADE. The t-SNE projection of CLADE and non-CLADE groups according to the subjective symptoms shows that the CLADE group had a variety of subjective symptoms compared with the non-CLADE group (14 groups vs 4 groups, respectively). Finally, we created a heatmap in which the patterns in the 18 groups based on each subjective symptom were stratified by the t-SNE projection. Those 18 groups were subgrouped based on the OSDI as follows: ocular symptoms (OSDI questions 1-5), vision-related function (questions 6-9), and environmental triggers (questions 10-12).

Figure 3. Stratification of subjective symptoms of CLADE. (A) The number of subjective symptoms in the non-CLADE and CLADE groups. (B) The frequency of subjective symptoms of DED based on each OSDI questionnaire. (C) The treemap shows the difference in the frequency of subjective symptoms of DED (the percentage of subjective symptoms in the CLADE group minus the percentage of subjective symptoms in the non-CLADE group) based on each OSDI question. (D) The Pearson correlation coefficients between subjective symptoms and CLADE are shown in the heatmap as a color gradient. (E) t-SNE projection shows the groups of CLADE based on the subjective symptoms. (F) The heatmap shows the correlation between the groups made by t-SNE projection and each OSDI item. CLADE: contact lens–associated dry eye. DED: dry eye disease. OSDI: Ocular Surface Disease Index. t-SNE: t-distributed Stochastic Neighbor Embedding.



Risk Factors for CLADE

Table 2 shows univariate odds ratios of each risk factor for CLADE, determined by logistic regression analysis. The univariate odds ratios were 0.99 (95% CI 0.975-0.998) for age, 1.75 (95% CI 1.33-2.23) for female sex, 0.98 (95% CI 0.97-0.99) for height, 1.41 (95% CI 1.13-1.77) for hay fever, 3.13 (95% CI 1.43-6.84) for mental illness other than depression or schizophrenia, 2.11 (95% CI 1.58-2.82) for past diagnosis of DED, 1.56 (95% CI 1.11-2.19) for more than 8 hours of screen exposure per day, and 1.69 (95% CI 1.26-2.28) for smoking. Relevant risk factors for CLADE were identified by the analyses of univariate odds ratios and were further analyzed by multivariate analyses. Table 2 also shows the multivariate odds ratios of each risk factor for CLADE. The multivariate odds ratios were 0.98 (95% CI 0.96-0.99) for age, 1.53 (95% CI 1.05-2.24) for female sex, 1.38 (95% CI 1.10-1.74) for hay fever, 2.51 (95% CI 1.13-5.57) for mental disease other than depression or schizophrenia, 2.21 (95% CI 1.63-2.99) for past diagnosis of DED, 1.61 (95% CI 1.13-2.28) for more than 8 hours of screen exposure per day, and 2.07 (95% CI 1.49-2.88) for smoking. Number of years of CL use, duration worn per day, and CL type had no significant association with CLADE.

Table 2. Univariate and multivariate adjusted odds ratios of risk factors for contact lens–associated dry eye.

Risk factors	Univariate odds ratio (95% CI)	Multivariate odds ratio (95% CI)
Demographic characteristics		
Age (per year)	0.99 (0.975-0.998)	0.98 (0.96-0.99)
Sex (female vs male)	1.75 (1.33-2.23)	1.53 (1.05-2.24)
Height (cm)	0.98 (0.97-0.99)	1.00 (0.98-1.02)
Weight (kg)	0.99 (0.98-1.00)	N/A[a]
CL[b] use (0-10 years)	0.98 (0.95-1.02)	0.99 (0.95-1.03)
CL duration (hours per day)		
0-6	1 (N/A)	1 (N/A)
7-12	1.25 (0.77-2.05)	1.28 (0.76-2.14)
13-18	1.19 (0.74-1.92)	1.24 (0.74-2.06)
19-24	1.77 (0.93-3.36)	1.45 (0.73-2.85)
Contact lens type		
Hard	1 (N/A)	1 (N/A)
Disposable (daily)	0.98 (0.61-1.58)	0.82 (0.49-1.36)
Disposable (biweekly[c])	0.87 (0.54-1.40)	0.76 (0.46-1.28)
Disposable (monthly)	1.30 (0.71-2.39)	0.92 (0.48-1.77)
Conventional lens	1.23 (0.63-2.39)	0.90 (0.44-1.83)
Medicated hypertension (yes vs no)	0.51 (0.23-1.16)	N/A
Diabetes (yes vs no)	0.96 (0.20-4.64)	N/A
Systemic diseases (yes vs no)		
Blood disease	1.92 (0.24-15.69)	N/A
Brain disease	0.68 (0.13-3.54)	N/A
Collagen disease	2.20 (0.27-17.65)	N/A
Heart disease	2.12 (0.63-7.10)	N/A
Kidney disease	2.91 (0.68-12.50)	N/A
Liver disease	1.51 (0.33-6.85)	N/A
Malignant tumor	N/A	N/A
Respiratory disease	1.62 (0.96-2.75)	N/A
Hay fever (yes vs no)	1.41 (1.13-1.77)	1.38 (1.10-1.74)
Mental illness (yes vs no)		
Depression	1.91 (0.86-4.26)	N/A
Schizophrenia	1.01 (0.28-3.62)	N/A
Other	3.13 (1.43-6.84)	2.51 (1.13-5.57)
Past diagnosis of dry eye disease	2.11 (1.58-2.82)	2.21 (1.63-2.99)
Ophthalmic surgery (yes vs no)		
Cataract surgery	N/A	N/A
LASIK[d]	0.27 (0.04-1.95)	N/A
Other	2.58 (0.91-7.28)	N/A
Coffee (per cup per day)	1.00 (0.92-1.10)	N/A
Screen exposure time (hours)		

Risk factors	Univariate odds ratio (95% CI)	Multivariate odds ratio (95% CI)
<4	1 (N/A)	1 (N/A)
4-8	1.08 (0.80-1.45)	1.07 (0.79-1.45)
>8	1.56 (1.11-2.19)	1.61 (1.13-2.28)
Periodic exercise (yes vs no)	0.83 (0.66-1.05)	N/A
Sleeping time (per hour per day)		
<6	1.05 (0.82-1.36)	N/A
6-9	1 (N/A)	N/A
>9	1.24 (0.91-1.68)	N/A
Smoking (yes vs no)	1.69 (1.26-2.28)	2.07 (1.49-2.88)
Water intake (per 100 mL/d)	1.00 (1.00-1.00)	N/A

[a] N/A: not applicable.

[b] CL: contact lens.

[c] Biweekly: every 2 weeks.

[d] LASIK: laser-assisted in situ keratomileusis.

Discussion

CL use significantly contributes to the quality of vision by correcting refractive errors. However, many CL users discontinue CL wear because of CLD [1-6]. This study analyzed individuals' medical data obtained from the DryEyeRhythm app and clarified the characteristics of subjective symptoms of CLADE and its risk factors. We found that CLADE remained undiagnosed in many individuals who experienced dry eye symptoms and had not been treated with eye drops. Since CLADE is a risk factor for CL discontinuation, evaluating individual subjective symptoms of CLADE and identifying personalized and preemptive medical care will contribute to more comfortable CL use. This study found that younger age, female sex, hay fever, mental illness other than depression or schizophrenia, extended screen exposure time, and smoking were CLADE risk factors. Our findings might help develop individual preemptive strategies for CLADE.

New medical big data collected from mobile health apps and the Internet of Medical Things have been used in recent years [18,22,35-37]. Because CL users are relatively young [38,39], using innovative methods such as smartphone apps is crucial in investigating individual subjective symptoms. This app recruitment model is more inclusive of younger people because they are relatively healthy and seldom visit hospitals. Indeed, many younger individuals (median age of 23 years) participated in this crowdsourced study due to DryEyeRhythm. Previous studies presented challenges regarding information collection for individuals who previously wore CL [1]; however, this crowdsourced study allowed an easier collection of the information compared with the conventional hospital-based study. This study found that 41.40% (1844/4454) of current CL users and 10.42% (464/4454) of past CL users discontinued CL use for various reasons, as Multimedia Appendix 4 showed. Moreover, as demonstrated in previous studies [40], the proportion of current CL users tended to decrease with increasing age, as seen in Multimedia Media Appendix 4. Our findings are consistent with those of previous studies using hospital-based, mail-based, email-based, and Facebook-based methods, which found that between 12% and 51% of CL users discontinued use [2,3,6]. Therefore, this mobile-based health study can be used to supplement traditional hospital-based studies.

Because 23% of the symptomatic participants did not exhibit typical clinical signs of dryness [41], investigating subjective symptoms of CLADE is likely to have more diagnostic value than conducting clinical tests. In this study, 78.47% (1447/1844) of current CL users had CLADE with an OSDI score ≥13. Given that only 28.82% (417/1447) of participants with CLADE had been diagnosed with DED in the past and only 27.92% (404/1447) of participants with CLADE had used eye drops, this study could assess the proportion of individuals who were undiagnosed with DED and did not undergo treatment intervention while experiencing CLADE symptoms. However, since CLADE presents various subjective symptoms, it is possible that those CL users might already have been experiencing CLADE symptoms but were not diagnosed. This study found that individuals with CLADE had multiple subjective symptoms, and more than 90% of participants with CLADE reported that their eyes felt gritty (1355/1447 participants, 93.64%) and were sensitive to light (1328/1447 participants, 91.78%). In particular, items related to environmental triggers were more frequent in CLADE patients than in non-CLADE patients (Figure 3), indicating that CLADE may be strongly influenced by environmental factors. Therefore, improvement of environmental triggers may be a potential intervention method to prevent CL discontinuation due to CLADE. Furthermore, this study stratified various individual subjective symptoms of CLADE using a multidimensional analysis with t-SNE (Figure 3), and the subjective symptoms of CLADE were divided into 14 subgroups. Some CLADE subgroups were strongly related to environmental factors and others were not. The findings indicate that it is important to conduct personalized medicine based on individual CLADE symptoms.

Our study aimed to identify risk factors that contribute to CLADE in a large-scale prospective clinical study using real-world data. The resulting data are shown as odds ratios of risk factors for CLADE (Table 2), including younger age, female sex, hay fever, mental illness other than depression or schizophrenia, past diagnosis of DED, extended screen exposure time, and smoking. Among these risk factors, female sex, mental illness other than depression or schizophrenia, extended screen exposure time, and smoking are associated with DED [17]. Young age is also a risk factor for CLADE, probably because of the higher sensitivity to CLADE symptoms among CL users in the younger cohort and the cessation of CL use in the older cohort due to DED [11,24]. Our study revealed that CLADE was recognized by many young CL users, suggesting the importance of the prevention and treatment of CLADE among young CL users. It should be noted that many CL users were women (1471/1844, 79.77%), and although our results do not directly associate the physiology of female sex with the pathology of CLADE, we believe that the significantly higher prevalence of CL use in the female cohort warrants the recognition of the female population as a relative risk factor for CLADE and DED. This study showed that the median wearing time per day of CL was 14 hours (IQR 12-16), indicating that CL are worn almost all day. We also demonstrated that over 8 hours of screen exposure time was a risk factor for CLADE, and many of the individuals in this study had more than 8 hours of screen exposure time. Therefore, to improve CLADE symptoms, it is necessary to propose a limit on screen time while wearing CL. Additionally, recent studies have demonstrated that hay fever and DED are pathologically related [42,43], thereby positing a synergistic effect of hay fever and DED on exacerbating CLADE symptoms. Moreover, our previous study demonstrated that hay fever, extended screen exposure time, and smoking are risk factors for severe dry eye symptoms [18]. Notably, these risk factors are modifiable and can be improved by lifestyle management [17]. Our findings would help identify individuals who are not yet diagnosed with CLADE and prevent deterioration of CLADE in routine clinical service and life.

Additionally, the types of CL were also identified using real-world data. Most of the CL users wore SCL and disposable lenses (Multimedia Appendix 5). However, the CL type and the daily CL duration did not correlate with CLADE (Table 2), as demonstrated in previous studies [44-46]. This study is a cross-sectional observational study; the causal relationship between CL type or daily duration of CL use and CLADE or CL discontinuation cannot be determined. Therefore, further study is needed to elucidate the associations between them.

Our study has several limitations associated with crowdsourced research, as presented in our previous studies [18,24,25]. First, this crowdsourced clinical study was characterized by selection bias for age, socioeconomic factors, and user characteristics because this app was released only for iOS (Apple Inc) devices. Furthermore, participants who were more interested in CLADE and had experienced CLADE symptoms might have completed all surveys and were subsequently included (Multimedia Appendix 3). Therefore, the prevalence of CLADE might have been overestimated. Second, this study might have recall bias because this study employed many self-administered questionnaires. We demonstrated the internal validity of the study [18]; however, considering the health-seeking behavior and cultural factors in Japan, the external validity or generalizability of the findings remains unknown. Additionally, socioeconomic status, education level, cultural background, and some important unmeasured factors related to CLADE were not investigated. In particular, this study found that one of the risk factors for CLADE was mental illness other than depression and schizophrenia, indicating that further precise classification for mental illness is needed. Further updates and development of an Android version of DryEyeRhythm and recruitment of individuals from other countries will reduce these biases. Third, this was a cross-sectional study; therefore, temporal relationships and causality between CLADE and the risk factors cannot be inferred. Additionally, this mobile health app study identified symptomatic dry eye based only on the OSDI questionnaire and might contain false-positives because dry eye examinations, including the Schirmer's test and measurement of tear film break-up time, were not performed. However, this crowdsourced clinical study using DryEyeRhythm overcame several common participant recruitment–related issues, including diverse cohort and geographic restrictions, thus leading to the collection of real-world medical big data. Notably, it would be difficult to identify non-CLADE without our mobile app. Moreover, DryEyeRhythm also presents a unique opportunity for preventive care by identifying individuals at risk for CLADE earlier than currently possible. The app can be used to supplement traditional hospital-based research, thereby encompassing a broader population.

In conclusion, we identified individuals with CLADE and the associated risk factors using DryEyeRhythm. The various subjective symptoms of CLADE collected and stratified in this study could be used for the future prevention of CLADE.

Acknowledgments

Special thanks to Ohako Inc for developing the DryEyeRhythm app, and Tina Shiang, Yosuke Yoshimura, Yoshimune Hiratsuka, and Miki Uchino for the initial development of the app. This study was supported by Seed Co, Ltd; Alcon Japan, Ltd; Rohto Pharmaceutical Co, Ltd; Hoya Corporation; and Wakamoto Co, Ltd. The sponsors had no role in the design or conduct of this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.
[DOC File , 76 KB - jmir_v22i6e18996_app1.doc]

Multimedia Appendix 2
Daily subjective symptoms questions.
[DOC File , 58 KB - jmir_v22i6e18996_app2.doc]

Multimedia Appendix 3
Sensitivity analysis between the included and excluded participants.
[DOCX File , 28 KB - jmir_v22i6e18996_app3.docx]

Multimedia Appendix 4
Characteristics of participants.
[DOCX File , 42 KB - jmir_v22i6e18996_app4.docx]

Multimedia Appendix 5
Characteristics of current contact lens users.
[DOCX File , 43 KB - jmir_v22i6e18996_app5.docx]

Multimedia Appendix 6
Age distributions of CL users.
[DOC File , 150 KB - jmir_v22i6e18996_app6.doc]

References

1. Dumbleton K, Caffery B, Dogru M, Hickson-Curran S, Kern J, Kojima T, members of the TFOS International Workshop on Contact Lens Discomfort. The TFOS International Workshop on Contact Lens Discomfort: report of the subcommittee on epidemiology. Invest Ophthalmol Vis Sci 2013 Oct 18;54(11):TFOS20-TFOS36. [doi: 10.1167/iovs.13-13125] [Medline: 24058130]
2. Pritchard N, Fonn D, Brazeau D. Discontinuation of contact lens wear: a survey. Int Contact Lens Clin 1999 Nov;26(6):157-162. [doi: 10.1016/s0892-8967(01)00040-2] [Medline: 11384832]
3. Dumbleton K, Woods CA, Jones LW, Fonn D. The impact of contemporary contact lenses on contact lens discontinuation. Eye Contact Lens 2013 Jan;39(1):93-99. [doi: 10.1097/ICL.0b013e318271caf4] [Medline: 23266586]
4. Wirth RJ, Edwards MC, Henderson M, Henderson T, Olivares G, Houts CR. Development of the Contact Lens User Experience: CLUE Scales. Optom Vis Sci 2016 Aug;93(8):801-808 [FREE Full text] [doi: 10.1097/OPX.0000000000000913] [Medline: 27383257]
5. Sindt CW, Longmuir RA. Contact lens strategies for the patient with dry eye. Ocul Surf 2007 Oct;5(4):294-307. [doi: 10.1016/s1542-0124(12)70095-2] [Medline: 17938839]
6. Richdale K, Sinnott LT, Skadahl E, Nichols JJ. Frequency of and factors associated with contact lens dissatisfaction and discontinuation. Cornea 2007 Feb;26(2):168-174. [doi: 10.1097/01.ico.0000248382.32143.86] [Medline: 17251807]
7. Tsubota K, Yokoi N, Shimazaki J, Watanabe H, Dogru M, Yamada M, Asia Dry Eye Society. New Perspectives on Dry Eye Definition and Diagnosis: A Consensus Report by the Asia Dry Eye Society. Ocul Surf 2017 Jan;15(1):65-76 [FREE Full text] [doi: 10.1016/j.jtos.2016.09.003] [Medline: 27725302]
8. Craig JP, Nichols KK, Akpek EK, Caffery B, Dua HS, Joo C, et al. TFOS DEWS II Definition and Classification Report. Ocul Surf 2017 Jul;15(3):276-283. [doi: 10.1016/j.jtos.2017.05.008] [Medline: 28736335]
9. Gayton JL. Etiology, prevalence, and treatment of dry eye disease. Clin Ophthalmol 2009;3:405-412 [FREE Full text] [doi: 10.2147/opth.s5555] [Medline: 19688028]
10. Shimmura S, Shimazaki J, Tsubota K. Results of a population-based questionnaire on the symptoms and lifestyles associated with dry eye. Cornea 1999 Jul;18(4):408-411. [doi: 10.1097/00003226-199907000-00003] [Medline: 10422851]
11. Ding J, Sullivan DA. Aging and dry eye disease. Exp Gerontol 2012 Jul;47(7):483-490 [FREE Full text] [doi: 10.1016/j.exger.2012.03.020] [Medline: 22569356]
12. Courtin R, Pereira B, Naughton G, Chamoux A, Chiambaretta F, Lanhers C, et al. Prevalence of dry eye disease in visual display terminal workers: a systematic review and meta-analysis. BMJ Open 2016 Jan 14;6(1):e009675 [FREE Full text] [doi: 10.1136/bmjopen-2015-009675] [Medline: 26769784]
13. Inomata T, Shiang T, Iwagami M, Sakemi F, Fujimoto K, Okumura Y, et al. Changes in Distribution of Dry Eye Disease by the New 2016 Diagnostic Criteria from the Asia Dry Eye Society. Sci Rep 2018 Jan 30;8(1):1918 [FREE Full text] [doi: 10.1038/s41598-018-19775-3] [Medline: 29382858]

14. du Toit R, Situ P, Simpson T, Fonn D. The effects of six months of contact lens wear on the tear film, ocular surfaces, and symptoms of presbyopes. Optom Vis Sci 2001 Jun;78(6):455-462. [doi: 10.1097/00006324-200106000-00020] [Medline: 11444637]
15. Chalmers RL, Begley CG. Dryness symptoms among an unselected clinical population with and without contact lens wear. Cont Lens Anterior Eye 2006 Mar;29(1):25-30. [doi: 10.1016/j.clae.2005.12.004] [Medline: 16448840]
16. Begley CG, Chalmers RL, Mitchell GL, Nichols KK, Caffery B, Simpson T, et al. Characterization of ocular surface symptoms from optometric practices in North America. Cornea 2001 Aug;20(6):610-618. [doi: 10.1097/00003226-200108000-00011] [Medline: 11473162]
17. Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, Malet F, et al. TFOS DEWS II Epidemiology Report. The Ocular Surface 2017 Jul;15(3):334-365. [doi: 10.1016/j.jtos.2017.05.003]
18. Inomata T, Nakamura M, Iwagami M, Shiang T, Yoshimura Y, Fujimoto K, et al. Risk Factors for Severe Dry Eye Disease: Crowdsourced Research Using DryEyeRhythm. Ophthalmology 2019 May;126(5):766-768. [doi: 10.1016/j.ophtha.2018.12.013] [Medline: 30550734]
19. Nichols JJ, Sinnott LT. Tear film, contact lens, and patient-related factors associated with contact lens-related dry eye. Invest Ophthalmol Vis Sci 2006 Apr;47(4):1319-1328. [doi: 10.1167/iovs.05-1392] [Medline: 16565363]
20. Chong PQY, Yeo S, Too CL, Boo C, Tong L. Effects of wearing a daily disposable lens on tear film: a randomised controlled trial. Clin Exp Optom 2016 May;99(3):241-247 [FREE Full text] [doi: 10.1111/cxo.12357] [Medline: 27087449]
21. Kojima T, Ibrahim OMA, Wakamatsu T, Tsuyama A, Ogawa J, Matsumoto Y, et al. The impact of contact lens wear and visual display terminal work on ocular surface and tear functions in office workers. Am J Ophthalmol 2011 Dec;152(6):933-940.e2. [doi: 10.1016/j.ajo.2011.05.025] [Medline: 21871600]
22. Heidari M, Noorizadeh F, Wu K, Inomata T, Mashaghi A. Dry Eye Disease: Emerging Approaches to Disease Analysis and Therapy. J Clin Med 2019 Sep 11;8(9):01439 [FREE Full text] [doi: 10.3390/jcm8091439] [Medline: 31514344]
23. Inomata T, Iwagami M, Hiratsuka Y, Fujimoto K, Okumura Y, Shiang T, et al. Maximum blink interval is associated with tear film breakup time: A new simple, screening test for dry eye disease. Sci Rep 2018 Sep 07;8(1):13443 [FREE Full text] [doi: 10.1038/s41598-018-31814-7] [Medline: 30194447]
24. Inomata T, Iwagami M, Nakamura M, Shiang T, Yoshimura Y, Fujimoto K, et al. Characteristics and Risk Factors Associated With Diagnosed and Undiagnosed Symptomatic Dry Eye Using a Smartphone Application. JAMA Ophthalmol 2019 Nov 27:58-68. [doi: 10.1001/jamaophthalmol.2019.4815] [Medline: 31774457]
25. Inomata T, Iwagami M, Nakamura M, Shiang T, Fujimoto K, Okumura Y, et al. Association between dry eye and depressive symptoms: Large-scale crowdsourced research using the DryEyeRhythm iPhone application. Ocul Surf 2020 Apr;18(2):312-319. [doi: 10.1016/j.jtos.2020.02.007] [Medline: 32113987]
26. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
27. Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the Ocular Surface Disease Index. Arch Ophthalmol 2000 May;118(5):615-621. [doi: 10.1001/archopht.118.5.615] [Medline: 10815152]
28. Zung W. A self-rating depression scale. Arch Gen Psychiatry 1965 Jan;12:63-70. [doi: 10.1001/archpsyc.1965.01720310065008] [Medline: 14221692]
29. Midorikawa-Inomata A, Inomata T, Nojiri S, Nakamura M, Iwagami M, Fujimoto K, et al. Reliability and validity of the Japanese version of the Ocular Surface Disease Index for dry eye disease. BMJ Open 2019 Nov 25;9(11):e033940 [FREE Full text] [doi: 10.1136/bmjopen-2019-033940] [Medline: 31772113]
30. Miller KL, Walt JG, Mink DR, Satram-Hoang S, Wilson SE, Perry HD, et al. Minimal clinically important difference for the ocular surface disease index. Arch Ophthalmol 2010 Jan;128(1):94-101. [doi: 10.1001/archophthalmol.2009.356] [Medline: 20065224]
31. Fujimoto K, Inomata T, Okumura Y, Iwata N, Fujio K, Eguchi A, et al. Comparison of corneal thickness in patients with dry eye disease using the Pentacam rotating Scheimpflug camera and anterior segment optical coherence tomography. PLoS ONE 2020 Feb 3;15(2):e0228567. [doi: 10.1371/journal.pone.0228567]
32. Biggs JT, Wylie LT, Ziegler VE. Validity of the Zung Self-rating Depression Scale. Br J Psychiatry 1978 Apr;132:381-385. [doi: 10.1192/bjp.132.4.381] [Medline: 638392]
33. Kitamura T, Shima S, Sugawara M, Toda MA. Temporal variation of validity of self-rating questionnaires: repeated use of the General Health Questionnaire and Zung's Self-rating Depression Scale among women during antenatal and postnatal periods. Acta Psychiatr Scand 1994 Dec;90(6):446-450. [doi: 10.1111/j.1600-0447.1994.tb01622.x] [Medline: 7892778]
34. Maaten L. Accelerating t-SNE using Tree-Based Algorithms. The Journal of Machine Learning Research 2014/1/1 2014;15(1):3221-3245.
35. Basatneh R, Najafi B, Armstrong DG. Health Sensors, Smart Home Devices, and the Internet of Medical Things: An Opportunity for Dramatic Improvement in Care for the Lower Extremity Complications of Diabetes. J Diabetes Sci Technol 2018 May;12(3):577-586. [doi: 10.1177/1932296818768618] [Medline: 29635931]
36. Armstrong DG, Najafi B, Shahinpoor M. Potential Applications of Smart Multifunctional Wearable Materials to Gerontology. Gerontology 2017;63(3):287-298. [doi: 10.1159/000455011] [Medline: 28076861]

37. Miller JD, Najafi B, Armstrong DG. Current Standards and Advances in Diabetic Ulcer Prevention and Elderly Fall Prevention Using Wearable Technology. Curr Geri Rep 2015 Jul 9;4(3):249-256. [doi: 10.1007/s13670-015-0136-7]
38. Lee YC, Lim CW, Saw SM, Koh D. The prevalence and pattern of contact lens use in a Singapore community. CLAO J 2000 Jan;26(1):21-25. [Medline: 10656305]
39. Swanson MW. A cross-sectional analysis of U.S. contact lens user demographics. Optom Vis Sci 2012 Jun;89(6):839-848. [doi: 10.1097/OPX.0b013e318255da45] [Medline: 22544000]
40. Asbell PA. Contact Lens Discomfort: Can We Prevent Dropout? Eye Contact Lens 2017 Jan;43(1):1. [doi: 10.1097/ICL.0000000000000358] [Medline: 28002222]
41. Young G, Chalmers R, Napier L, Kern J, Hunt C, Dumbleton K. Soft contact lens-related dryness with and without clinical signs. Optom Vis Sci 2012 Aug;89(8):1125-1132. [doi: 10.1097/OPX.0b013e3182640af8] [Medline: 22820475]
42. Ayaki M, Kawashima M, Uchino M, Tsubota K, Negishi K. Possible association between subtypes of dry eye disease and seasonal variation. Clin Ophthalmol 2017;11:1769-1775 [FREE Full text] [doi: 10.2147/OPTH.S148650] [Medline: 29033539]
43. Fujishima H, Toda I, Shimazaki J, Tsubota K. Allergic conjunctivitis and dry eye. Br J Ophthalmol 1996 Nov;80(11):994-997 [FREE Full text] [doi: 10.1136/bjo.80.11.994] [Medline: 8976728]
44. Lubis RR, Gultom MTH. The Correlation between Daily Lens Wear Duration and Dry Eye Syndrome. Open Access Maced J Med Sci 2018 May 20;6(5):829-834 [FREE Full text] [doi: 10.3889/oamjms.2018.215] [Medline: 29875854]
45. Sapkota K, Martin R, Franco S, Lira M. Common symptoms of Nepalese soft contact lens wearers: A pilot study. J Optom 2015;8(3):200-205 [FREE Full text] [doi: 10.1016/j.optom.2015.01.004] [Medline: 25659748]
46. Young G, Chalmers RL, Napier L, Hunt C, Kern J. Characterizing contact lens-related dryness symptoms in a cross-section of UK soft lens wearers. Cont Lens Anterior Eye 2011 Apr;34(2):64-70. [doi: 10.1016/j.clae.2010.08.005] [Medline: 20832350]

Abbreviations

CL: contact lens
CLD: contact lens discomfort
CLADE: contact lens–associated dry eye
DED: dry eye disease
OSDI: Ocular Surface Disease Index
SDS: Zung Self-Rating Depression Scale
t-SNE: t-distributed Stochastic Neighbor Embedding

Edited by G Eysenbach; submitted 31.03.20; peer-reviewed by K Co Shih, T Yamaguchi; comments to author 18.04.20; revised version received 23.04.20; accepted 27.04.20; published 26.06.20.

Please cite as:
Inomata T, Nakamura M, Iwagami M, Midorikawa-Inomata A, Sung J, Fujimoto K, Okumura Y, Eguchi A, Iwata N, Miura M, Fujio K, Nagino K, Hori S, Tsubota K, Dana R, Murakami A
Stratification of Individual Symptoms of Contact Lens–Associated Dry Eye Using the iPhone App DryEyeRhythm: Crowdsourced Cross-Sectional Study
J Med Internet Res 2020;22(6):e18996
URL: http://www.jmir.org/2020/6/e18996/
doi:10.2196/18996
PMID:32589162

Original Paper

Challenges and Successes in Raising a Child With Type 1 Diabetes and Autism Spectrum Disorder: Mixed Methods Study

Tamara K Oser[1], MD; Sean M Oser[1], MD, MPH; Jessica A Parascando[2], MPH; Lee Ann Grisolano[3], PhD; Kanthi Bangalore Krishna[4], MD; Daniel E Hale[4], MD; Michelle Litchman[5], PhD, FNP-BC; Shideh Majidi[6], MD; Paul Haidet[7], MD, MPH

[1]Department of Family Medicine, University of Colorado School of Medicine, Aurora, CO, United States
[2]Department of Family and Community Medicine, Penn State College of Medicine, Hershey, PA, United States
[3]Department of Psychiatry, Penn State College of Medicine, Herhsey, PA, United States
[4]Division of Pediatric Endocrinology, Department of Pediatrics, Penn State of College of Medicine, Hershey, PA, United States
[5]College of Nursing, University of Utah, Salt Lake City, UT, United States
[6]Barbara Davis Center, University of Colorado School of Medicine, Aurora, CO, United States
[7]Departments of Medicine, Public Health Sciences, and Humanities, Penn State College of Medicine, Hershey, PA, United States

Corresponding Author:
Tamara K Oser, MD
Department of Family Medicine
University of Colorado School of Medicine
12631 East 17th Avenue, F496
Aurora, CO, 80045
United States
Phone: 1 3037249700
Email: tamara.oser@cuanschutz.edu

Abstract

Background: Self-management of type 1 diabetes (T1D) requires numerous decisions and actions by people with T1D and their caregivers and poses many daily challenges. For those with T1D and a developmental disorder such as autism spectrum disorder (ASD), more complex challenges arise, though these remain largely unstudied.

Objective: This study aimed to better understand the barriers and facilitators of raising a child with T1D and ASD. Secondary analysis of web-based content (phase 1) and telephone interviews (phase 2) were conducted to further expand the existing knowledge on the challenges and successes faced by these families.

Methods: Phase 1 involved a qualitative analysis of publicly available online forums and blog posts by caregivers of children with both T1D and ASD. Themes from phase 1 were used to create an interview guide for further in-depth exploration via interviews. In phase 2, caregivers of children with both T1D and ASD were recruited from Penn State Health endocrinology clinics and through the web from social media posts to T1D-focused groups and sites. Interested respondents were directed to a secure web-based eligibility assessment. Information related to T1D and ASD diagnosis, contact information, and demographics were collected. On the basis of survey responses, participants were selected for a follow-up telephone interview and were asked to complete the adaptive behavior assessment system, third edition parent form to assess autism severity and upload a copy of their child’s most recent hemoglobin A_{1c} (HbA_{1c}) result. Interviews were transcribed, imported into NVivo qualitative data management software, and analyzed to determine common themes related to barriers and facilitators of raising a child with both ASD and T1D.

Results: For phase 1, 398 forum posts and blog posts between 2009 and 2016 were analyzed. Common themes related to a lack of understanding by the separate ASD and T1D caregiver communities, advice on coping techniques, rules and routines, and descriptions of the health care experience. For phase 2, 12 eligible respondents were interviewed. For interviewees, the average age of the child at diagnosis with T1D and ASD was 7.92 years and 5.55 years, respectively. Average self-reported and documented HbA_{1c} levels for children with T1D and ASD were 8.6% (70 mmol/mol) and 8.7% (72 mmol/mol), respectively. Common themes from the interviews related to increased emotional burden, frustration surrounding the amount of information they are expected to learn, and challenges in the school setting.

Conclusions: Caregivers of children with both T1D and ASD face unique challenges, distinct from those faced by caregivers of individuals who have either disorder alone. Understanding these challenges may help health care providers in caring for this unique population. Referral to the diabetes online community may be a potential resource to supplement the care received by the medical community.

(J Med Internet Res 2020;22(6):e17184) doi:10.2196/17184

KEYWORDS

type 1 diabetes; autism spectrum disorder; child; blogs; social media; qualitative research

Introduction

Background

Individuals with type 1 diabetes (T1D) spend over 8000 hours per year self-managing diabetes outside of the medical setting [1]. Optimal management includes numerous daily behaviors and decisions such as checking blood glucose levels, counting carbohydrates, calculating doses of and administering insulin, and changing dietary habits. Following this regimen is challenging for anyone, but for people with developmental disorders, such as autism spectrum disorder (ASD), this complex care can be even more demanding. Individuals with ASD can have impairments in social communication and interaction as well as repetitive or restrictive behaviors and preferences [2]. Behaviors and activities that individuals with ASD perceive as unimportant or do not understand, such as self-care, health and safety awareness, and activities of daily living, can be very difficult for them to manage independently, which can make the complexity of self-managing diabetes even more challenging. In addition, over 30% of people with ASD also have an intellectual disability (ID), which can add additional challenges in the self-management of T1D [3].

T1D prevalence is increasing, affecting 1.25 million people in the United States in 2014, and 5 million people are expected to be affected by 2050 [4-6]. Up to 1 in 59 children in the United States have been identified with ASD [7], and the reported prevalence of ASD is also increasing [3,7]. There is little research regarding the prevalence of individuals with both ASD and T1D in the United States [3,6-8], although a recent study found the prevalence of ASD in the pediatric T1D population in 1 clinic to be 1.16% (95% CI 0.96-1.26) [9]. An additional study reported that among 10,032 participants from the T1D Exchange Clinic Network, 159 (1.58%) had T1D and ASD [10]. These studies suggest that there is no increased prevalence of ASD in children with diabetes. However, research has documented that people with ID alone may be more likely to develop T1D as a result of having genetic conditions such as Down syndrome [11,12]. In addition, 31% of children with ASD also have an ID [3]. Although ASD alone may not lead to increased diabetic ketoacidosis, severe hypoglycemia, or worsening glycemic control [10], T1D patients with both ASD and comorbid ID are more likely to have very poor glycemic control, higher emergency room utilization, and lower rates of achieving national standards of diabetes management, including lower rates of lipid screening and retinopathy screening [11-13]. With this comes greater risk for negative outcomes, and indeed diabetes-related morbidity and mortality have been identified as a particular concern for people with ID [14,15]. As people with ASD and ID live longer, it has become even more important that they achieve better long-term disease management, including glycemic control and control of other modifiable risk factors, to decrease the occurrence of potentially preventable complications as they age. The interplay between ASD with or without ID and impacts on clinical, behavioral, and psychosocial outcomes related to T1D continue to be areas where little has been explored. Many factors may be involved in such interplay and outcomes, including access to social support and social networks [16], which may be valuable resources in raising a child with both T1D and autism.

Although there have been great advances over the last decades in health care delivered to children with chronic conditions, not all children have benefited equally [17]. Although there is a well-developed evidence base regarding diabetes care and education for populations without ASD or ID, very little research has been done to better understand optimal ways to deliver education and behavioral interventions and to align clinical care with the needs of families raising a child with both T1D and ASD, with or without an ID.

Objectives

This study sought to better understand self-management challenges and successes experienced by families with experience in raising a child with both T1D and ASD to lay the groundwork for future studies. This will enable the exploration of interventions to improve diabetes outcomes and reduce health disparities in this special population and other unique populations of children living with diabetes and other chronic medical conditions. To the best of our knowledge, this is the first study that utilizes a qualitative component to allow further exploration of the lived experience of raising a child with both T1D and ASD. We also sought to gather some exploratory information on glycemic control related to the degree of adaptive function in light of the coexisting ASD, hypothesizing that adaptive impairment may be associated with hemoglobin A_{1c} (HbA_{1c}).

Methods

Design and Ethics

This study involved an innovative approach consisting of 2 phases and included the use of social media as well as traditional interviews to capitalize on the strengths of each. In the first phase, secondary data analysis of web-based content was conducted to understand the challenges and successes families raising a child with T1D and ASD experience, similar to our other studies using web-based content to evaluate barriers and facilitators of living with or raising a child with T1D [18,19]. The themes generated from this study were used to create an interview guide that was used in phase 2 to interview caregivers of children with varying degrees of glycemic control and ASD severity to provide richer exploration of the lived experience in raising a child with both T1D and ASD. This 2-step approach allowed further exploration of initial themes identified from the secondary data analysis of web-based content, with the advantage of being able to ask follow-up questions.

This study was approved by the Penn State College of Medicine institutional review board through expedited review and was considered to be *no greater than minimal risk.* No procedures or participant interactions took place before receiving this approval.

Phase 1

Recruitment

No individuals were actively recruited for the first phase of the study, as it consisted of a secondary data analysis of publicly available data. Content from a publicly available forum and 2 blogs were identified through an internet search using the search terms *type 1 diabetes and autism* and *diabetes and autism*, which would capture both *autism* and *autism spectrum disorder*.

Analysis

Qualitative methods are appropriate for generating new knowledge and hypotheses, especially when little is known about the area of investigation [20]. This type of analysis allows the emergence of themes from the data when there are no existing hypotheses to test, such as in this understudied population of children living with both T1D and ASD.

A total of 2 experienced qualitative researchers (TO and SO) undertook data analysis. Content from online forums and blogs were imported into NVivo version 12 qualitative software (QSR International Pty Ltd) post by post. After reviewing blogs and forum posts and noting initial impressions, a codebook was developed and revised through ongoing discussions among the study team. To establish Cohen κ (a measure of interrater reliability in qualitative coding) [21], the primary coders (TO and SO) each coded 20% of the dataset ($\kappa=0.996$). With the established interrater reliability, the remaining posts and blog comments in the full dataset were coded. The research team employed an exploratory inductive thematic approach to construct emergent themes [22]. Themes from the analysis of web-based content were used to create an interview guide for use in the second phase of the study.

Phase 2

Recruitment

For the second phase of the study, participants were recruited on the web, through physician referral, and with flyers posted in the waiting room of an academic pediatric endocrinology practice. Participants were considered eligible if they were a caregiver (parent, grandparent, adult sibling, or guardian) to a child aged 5-18 years with both T1D and ASD; able to read, write, and communicate effectively in English; and had telephone and internet access. A summary explanation of the research was provided, including implied consent upon reading the summary of explanation and agreeing to participate.

Instrument

The eligibility survey was hosted in Research Electronic Data Capture (REDCap), a secure electronic data collection system [23]. The survey was tested for technological function before distribution, and questions were answered on 1 page. REDCap automatically captures the date of survey completion and completeness of the survey and assigns each participant a unique study code number. Participants were unable to go back in the survey and change their answers. The survey could only be completed once per Internet Protocol address. Participants reported demographics and information related to their child’s conditions and diabetes provider.

Interview

Survey participants were also asked if they would consent to a subsequent 30-min telephone interview. The interview guide was developed based on the thematic qualitative analysis of web-based content in phase 1. This included themes that seemed fully or nearly fully contextualized by the web-based data as well as themes that may not have been as fully developed, as the phase 2 interviews were used to assess the potential completeness of the phase 1 themes as well as to elicit deeper and more nuanced exploration of the phase 1 themes. To do this, a random sample of interview participants from online recruitment sources was selected for interviews based on consent and responses to the survey questions. As certain demographic groups (eg, clinic patients) were underrepresented in this sample, we also purposively chose to interview the few clinic patients who completed the survey (n=2). Such a targeted, purposive approach to adding participants is standard in qualitative research, as it is 1 of the distinct features that increases efficiency and yield, thereby minimizing overrepresentation and unneeded effort from participants in groups adequately represented [24]. This strategy enabled an interview sample large enough to reach saturation in the interviews, as is also typical of such qualitative research [24].

Telephone interviews were scheduled based on the earliest convenience of the participant. The interviews lasted approximately 30 min and were conducted by 3 authors (PH, SO, and TO). Interviews were audio recorded on a hand-held device and then transferred to a secure web-based storage location with access limited to the study team and an approved Health Insurance Portability and Accountability Act (HIPAA)-compliant transcriptionist.

After scheduling the interview, participants received a link to the adaptive behavior assessment system, third edition (ABAS-3) parent form [25], also hosted in REDCap, and were instructed to complete it either before or after the telephone interview. The ABAS-3 parent form assesses the functional skills of children and adolescents who display various types of limitations, disabilities, or disorders, much like those subjectively displayed by children with ASD and ID. The items included *assess everyday activities required, eg, to function adaptively, respond to environmental demands, care for oneself, and interact with others*. The results of the ABAS-3 identify, among other things, strengths and limitations as well as the need for services and support [25,26]. The general adaptive composite (GAC) scores were derived from all of the adaptive skills areas in the assessment and ranged from 40 to 120. Higher scores indicate a greater level of daily functioning, whereas lower scores suggest impairments in any of a variety of adaptive areas that are assessed by the ABAS-3. The further the GAC score deviates downward from what is considered typical of an individual's age, the greater the impairment in function. For example, a GAC score of 65 would suggest a greater level of impairment in adaptive behavior than a GAC score of 95.

Participants were also asked to upload a copy of the child's most recent HbA_{1c} result (within the past 4 months) into the REDCap survey. After completing all procedures, participants were compensated for their time and effort with a gift card.

Analysis

In all, 12 respondents were interviewed. Interviews were transcribed using HIPAA-compliant Penn State transcription services. Interview transcripts were imported into NVivo 12 (QSR International) for analysis and analyzed as detailed above through an exploratory inductive thematic qualitative approach (Cohen κ after 20%=0.942). Attempts were made to include caregivers of children with varying degrees of glycemic control and varying degrees of function related to their ASD.

Descriptive statistics were calculated for ABAS-3 parent form data and for HbA_{1c} results were also calculated by subgroups based on HbA_{1c} being <7.5% (<58 mmol/mol) or ≥7.5% (≥58 mmol/mol) according to then-current American Diabetes Association glycemic target recommendations [27].

Results

Phase 1

There were 398 posts between 2009 and 2016 from a publicly available forum for caregivers raising a child with both T1D and ASD (117 unique members) and 25 individual blog posts from 2 blogs authored by a caregiver to a child with T1D and ASD. In all, 4 themes emerged from phase 1 (Table 1).

Table 1. Phase 1 themes and representative quotes from analysis of online forums and blogs.

Themes	Representative web-based quotes
Parents of children with T1D[a] and autism yearned for support and did not feel understood by the autism community or the T1D community; social media provided much needed peer support from families that understood the unique experience of raising a child with both conditions.	• "I feel left out of the T1D groups because Aspergers brings a whole new set of concerns. But the Aspie groups are completely afraid of T1D." • "Like you said—the autism community is not equipped for our extra needs and the diabetes community just doesn't get the autism side of it." • "I found this group less than a week ago, and WEPT tears of joy that there are other parents out there who know what this is like." • "I am so thrilled to have found this group! I knew there were other families like ours but I didn't know how/where to access them. I am in tears right now because I feel so overwhelmed knowing that there are other people who understand the everyday trials (and victories!) both autism and diabetes bring to your life 24/7."
Numerous coping techniques were discussed, including focusing on T1D management first, as glucose levels could affect behavior.	• "What has worked is focusing on diabetes management first. When his blood sugars are high he is so irritable and easily set off." • "We chose to focus on T1D first because I do believe that has a huge effect on his behavior."
Descriptions of the health care experience, including the helpfulness of multidisciplinary teams with endocrinologists and autism specialists jointly developing treatment plans.	• "But the Endocrinologists I've encountered don't seem to see it that way. Ours acts like autism and diabetes are as related as a sprained ankle + appendicitis. To them, diabetes is about numbers and autism doesn't affect those. (Ha!)" • "Psychiatrists have only a passing familiarity with insulin dependence from an hour or two during med school, and they are mostly only interested in how it intersects with the meds they prescribe. So again, you as the advocate have to navigate the separate systems, trying to educate whoever is willing to listen along the way." • "My suggestion would be that the Endo consult with your regular doctor. A treatment and medical management plan should be put into place between the two offices. You may also ask for nutritionist, OT and behavior consults to help your child learn more about his diabetes and how his anger issues play into his blood glucose levels."
Sensory issues precluding the use of technology such as insulin pumps and the utility of tubeless pumps in overcoming some of these challenges.	• "My son is on shots, BTW. We tried a pump for a brief time but he wouldn't tolerate it. He pulled it off up to three times a day." • "We went with the tubeless pump for the reasons that have come up...we were concerned about our son messing with the controls and we also knew that for a kid with sensory issues, the external tubing would be a real problem." • "There was NO WAY my son could have dealt with tubing, but he's not having problems with the tubeless pump, we've been very pleased." • "There is just no way my sensory son would have tolerated tubing so I am glad we pursued the tubeless pump option." • "The one thing that helped with sensory issues for things attached [insulin pump, CGM] was I started putting bandaids on for everything-- every shot, bump, bruise, etc. After a few years he hardly noticed them...and as long as the tubing is tucked in he won't use that as something to self-stimulate with."

[a] T1D: type 1 diabetes.

Phase 2

In phase 2, 12 participants were selected for a telephone interview from among 29 respondents who completed the eligibility survey. Information related to participant demographics and study variables such as HbA_{1c} and ABAS-3 parent form scores were obtained (Table 2). On average, both interview participants and the broader group of survey respondents were similar in current age, age of child's diagnosis with T1D, age of child's diagnosis with ASD, and self-reported HbA_{1c}. For both interview participants and survey respondents, the majority reported that they were white, not Hispanic or Latino, lived in a suburban setting, and visited their child's provider every 3-6 months. The majority of both groups reported using continuous glucose monitoring (CGM). The majority of survey respondents reported using an insulin pump, whereas interview participants were evenly split.

In the study sample, 8 individuals had the most recent HbA_{1c} >7.5%, with an average of 9.7% (SD 1.89), or 82.5 mmol/mol (SD 20.5); 4 individuals had the most recent HbA_{1c} <7.5%, averaging 6.9% (SD 0.29), or 52.8 mmol/mol (SD 2.9). There was no statistically significant difference in GAC scores between those with HbA_{1c} >7.5% (mean 68.6, SD 14.22) and those with HbA_{1c} <7.5% (mean 76.3, SD 11.73; t_{10}=0.92; *P*=.38).

Overall, 3 themes emerged from an inductive thematic analysis of content obtained from interviews during phase 2 (Table 3).

Table 2. Phase 2 participant characteristics.

Variable	Interview participants (n=12)	Survey respondents (not interviewed; n=17)
Child's current age (years), mean (SD)	12.3 (3.1)	12.6 (3.7)
Child's age at diagnosis with T1D[a] (years), mean (SD)	7.9 (2.8)	7.3 (3.4)
Child's age at diagnosis with ASD[b] (years), mean (SD)	5.6 (3.5)	5.6 (3.8)
Self-reported HbA_{1c}[c], % (mmol/mol)	8.6 (70)	7.7 (61)
Laboratory confirmed HbA_{1c}, % (mmol/mol)	8.7 (72)	N/A[d]
Country of residence, n (%)		
United States	12 (100)	17 (100)
Gender, n (%)		
Male	7 (58)	10 (59)
Female	5 (42)	7 (41)
Race, n (%)		
White	11 (92)	16 (94)
Black	0 (0)	1 (6)
More than 1 race	1 (8)	0 (0)
Ethnicity, n (%)		
Hispanic or Latino	1 (8)	0 (0)
Not Hispanic or Latino	11 (92)	16 (94)
Unknown/not reported	0 (0)	1 (6)
Setting, n (%)		
Suburban	10 (83)	10 (59)
Urban	2 (17)	4 (23)
Rural	0 (0)	3 (18)
CGM[e] use, n (%)		
Yes	8 (67)	13 (77)
No	4 (33)	4 (23)
Insulin pump use, n (%)		
Yes	6 (50)	14 (82)
No	6 (50)	3 (18)
Tubeless pump use among subset of pump users, n (%)	5 (83)	7 (50)
Distance to provider, n (%)		
0-10 miles	3 (25)	4 (23)
11-20 miles	2 (17)	3 (18)
21-50 miles	7 (58)	6 (35)
>50 miles	0 (0)	4 (23)
Frequency of provider visits (usual care is every 3 months), n (%)		
<3 months	1 (8)	0 (0)
3-6 months	11 (927)	16 (94)
>6 months	0 (0)	1 (6)
ABAS-3[f] GAC[g] standard score, mean (95% CI)	71.2 (67.9-74.4)	N/A

[a] T1D: type 1 diabetes.

[b] ASD: autism spectrum disorder.

[c] HbA_{1c}: hemoglobin A_{1c}.

[d] N/A: not applicable.

[e] CGM: continuous glucose monitoring.

[f] ABAS-3: adaptive behavior assessment system, third edition.

[g] GAC: general adaptive composite.

Table 3. Phase 2 themes and representative quotes from analysis of interviews.

Theme	Representative interview quotes
Caregivers of children with T1D[a] and autism face emotional burdens that may be more than additive compared with raising a child with only T1D or autism, and they describe the constant monitoring and work required to care for their children.	• “Autism doesn’t go away because of a diabetes diagnosis.” • “How are we going to do this forever?” • “I am a nervous wreck every day.” • “She doesn’t show any signs of being low until she is on the brink of losing consciousness and then all of a sudden she is almost passing out.” • “There is a different kind of fear when it comes to diabetes than with autism.” • “Things are very overwhelming---I cried myself to sleep a couple of nights last week.” • “We are working harder than any other parent in the diabetes clinic and no matter what we do or how much monitoring, or how much insulin we give, it’s like nothing seems to help.” • “We try so hard and it seems like it’s not doing anything.” • “He cannot tell me if his blood sugar is low or high so I always have to be right there, constantly monitoring.”
Caregivers of children with T1D and autism express frustration surrounding the amount of T1D information they are expected to learn and the manner in which it is presented.	• “Sometimes people should understand that the best learning is from somebody teaching you and not from reading it.” • One caregiver describes a nurse educator who told her “you should know how to do this already. Just change the insulin. You should know how to do that already,” and her concern “that’s a really dangerous statement to make to a parent who’s new, who you never taught to change those doses.” • “They showed me how to put it [the CGM] in and sent me on my merry way, which was the hardest thing ever to learn. I had to then put it in on my own and it took me three tries and I ruined two sensors. I felt so bad because I didn’t know how to do it.” • Another caregiver described how a health care provider “handed me the Pink Panther [reference book regarding type 1 diabetes] and thought I had read it all.” • “I feel like especially with the hurdles we deal with, a lot more education from the onset would be very helpful.” • Caregivers offered solutions such as “make step by step videos like the ‘how to’ videos seen on YouTube” and offering “extra appointments and longer appointments.”
Caregivers describe numerous challenges surrounding their child’s experience in the school setting.	• “I am literally on call even when she’s at school. Even though she has an aide and there’s a nurse, they will still call me to come help take care of her on a daily basis.” • “They’ve dealt with her eloping from the classroom for years and they know she’s a good hider. They knew about the diabetes. I mean, I know they’re only human, but my goodness. If you lose a diabetic autistic child, that’s kind of a big deal.” • “I stepped down from my position as a paralegal to take care of my child and I now work at my daughter’s school, just to make sure that she is ok.” • Multiple caregivers described how they had to take “time off”, “reduce to part-time,” or even “take jobs at the school to be available to assist with [their] child’s care”. • “Schools are not doing the best they can to help these kids. Sometimes they think the kids can do it on their own where you know they have more challenges and they need more help.” • “No one can work together to get us what we need.”

[a] T1D: type 1 diabetes.

Discussion

Principal Findings

Caregivers raising a child with comorbid T1D and ASD face unique challenges that are not well understood by families with a child with only 1 of the individual conditions or by health care providers, as evident by the themes that emerged from phase 1. Themes were generally related to the combination of T1D and ASD and focused on coping techniques, advice on needles and technology, health care experience, and sensory issues related to technology (eg, insulin pumps). Themes from phase 2 are complementary. Although some could be interpreted as applying to nearly anyone performing T1D management (with or without coexisting ASD), many of the diabetes-specific challenges were influenced (and generally amplified) by the coexistence of ASD.

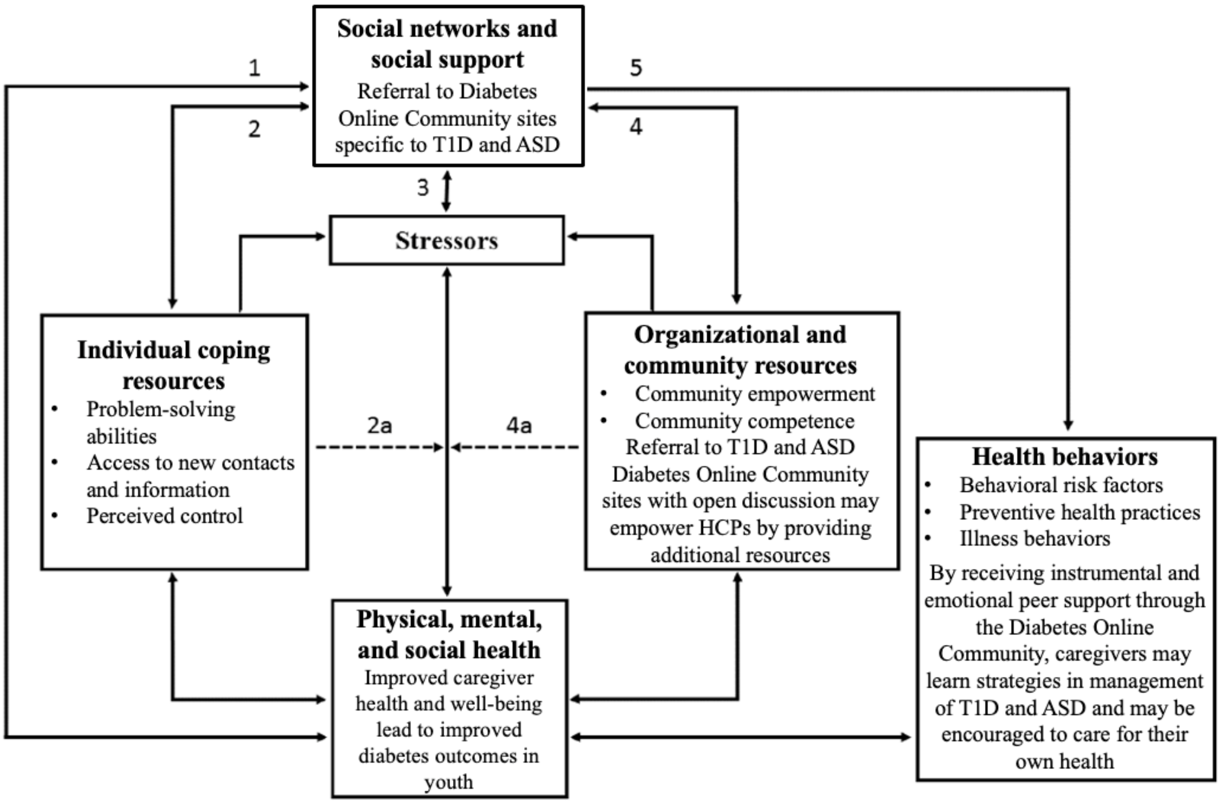
To the best of our knowledge, this is the first study with a large qualitative component to better understand the experience of raising a child with both T1D and ASD. The use of both social media and interviews as qualitative data sources in this special population provided a deeper understanding than using either type of data source alone, much like we found in another study using social media and interviews in a sample of adults with T1D [18].

Overall, our study results show that children who have greater impairment in function based on adaptive skills areas do not achieve target HbA_{1c} values. On the basis of the themes that emerged from our interviews (phase 2), this may be due to the additional barriers that come with the comorbidity, overwhelming amount of information to learn, and challenges in the school setting. More detailed comparisons around our exploratory aim were not conclusive due to the small sample size, not surprisingly.

Caregivers who shared the many challenges they face in raising a child with T1D and ASD might benefit from expanded social networks to provide social support, including emotional, instrumental, informational, and appraisal support. Both in-person support and referral to components of the online caregiver community specific to comorbid T1D and ASD could be valuable resources, as supportive ties may enhance well-being and health of the caregivers, which has been shown to lead to positive outcomes in children with T1D [28-34]. Social networks can provide much needed emotional support, exemplified by 1 participant who, upon finding such a web-based resource, “WEPT tears of joy that there are other parents out there who know what this is like.” Instrumental support might result if caregivers to a child with comorbid T1D and ASD find they are geographically able to provide assistance such as babysitting services, because finding caregivers for any child with T1D presents challenges, and these challenges may be significantly increased in a child with both T1D and ASD [19]. This study shows that caregivers identified numerous coping techniques, including focusing on T1D management first, establishing concrete rules and routines, advice on types of needles and technology, and suggestions for health care appointments. Prioritizing disease-specific tasks when diabetes coexists with other health conditions has been reported elsewhere [35]. Referral to social networks where this kind of informational support can be shared may be highly valuable. Appraisal support in the form of feedback may be especially useful to caregivers with a child newly diagnosed with either T1D or ASD in the setting of having previously been diagnosed with the other condition, as receiving feedback that the challenges they are facing are real and being handled to the best of the caregiver’s ability may help the caregiver in coping with the added new diagnosis.

A modified conceptual model for the relationship of social networks and social support to health can be applied to this unique population (Figure 1, adapted from Glanz et al [16]). Pathway 1 shows that through the expansion of social networks, the supportive ties developed may enhance caregiver health and well-being, which is linked in multiple studies to improved diabetes outcomes in their youth. Pathway 2 represents the hypothesized effect of social networks and social support on individuals’ coping resources. By being able to access new contacts and information to help solve the unique challenges faced by families raising a child with T1D and ASD, there may be a sense or perception of increased control that may also improve the physical and mental health of caregivers. The community of health care providers caring for these children should consider referral to web-based caregiver sites specific for families raising children with both T1D and ASD. By providing resources to these families to strengthen their social networks, the community of health care providers caring for these children may also feel more empowered in caring for this challenging population, as shown in pathway 4. Pathways 2a and 4a demonstrate the *buffering effect*, whereby the negative effects of the stressful experience of raising a child with T1D and ASD described in this study may be diminished through an expanded social network and having health care providers who understand the potential role of peer support in coping. Pathway 3 illustrates how receiving information through an expanded social network may reduce the amount of stress experienced, which, in turn, can lead to improved mental and physical health in caregivers of children with T1D and ASD. Finally, pathway 5 describes the potential effects of social networks and support on health behaviors. By exchanging information and support, caregivers to a child with T1D and ASD may be more supported in health behaviors that improve the care of their child and themselves as well.

Figure 1. A modified conceptual model for the relationship of social networks and social support to health, specific to type 1 diabetes and autism spectrum disorder. ASD: autism spectrum disorder; HCP: health care practitioner; T1D: type 1 diabetes.



Comparison With Previous Work

Our study demonstrated that the majority of survey respondents (8/12, 67%) and interview participants (13/17, 77%) had a child who used CGM technology. However, 82% (14/17) of the survey respondents reported their child using an insulin pump, compared with only 50% (6/12) of interview participants. Among those using insulin pumps, the majority use tubeless pumps (12/20, 60% overall, 5/6, 83% of interviewees and 7/14, 50% of survey respondents).

These findings complement the work of Stanek et al [9] and Bethin et al [10], who found that CGM use was the same among children with T1D and ASD as it was among those with T1D but without ASD, whereas insulin pump use was lower for those with both T1D and ASD. This qualitative work adds some explanatory context, as caregivers identified that sensory issues can preclude the successful use of insulin pumps and that successful solutions for some have included the use of tubeless pumps and desensitization techniques, such as the routine use of band-aids before attempting to use diabetes technology.

HbA_{1c} among study participants was not statistically correlated with the degree of function related to ASD, but it was limited by sample size. However, the trends of 2 comparisons (mean GAC scores of those achieving target HbA_{1c} vs those not achieving target HbA_{1c} and correlation of GAC scores with HbA_{1c}) were both in the direction that one would expect: children who scored higher in skills of daily living had improved HbA_{1c} values (<7.5%). It is not surprising that this study did not achieve statistical significance, as it was not powered to do so and the sample size was chosen for the greater qualitative focus here. This also echoes comments by Bethin et al [10] and Stanek et al [9] that further work is needed to be able to better support these children and their families. From this study specifically, we propose that further research with larger sample sizes is needed to continue to explore the potential relationships between the degree of functioning related to ASD and glycemic control.

Limitations

Limitations of this study include the lack of representation of caregivers from diverse ethnic backgrounds and rural areas. In addition, although attempts were made to recruit both from an endocrinology clinic and the diabetes online community (DOC), a minority of participants were recruited through the clinic. An additional limitation is that the search strategy included *autism* to capture ASDs but did not specifically include *Asperger* and may, therefore, have limited the scope of search returns. However, this would not fully exclude representation of Asperger's-related content as it is an ASD, and it would also tend to skew the sample toward more severe levels of impairment, increasing the ability to identify challenges within the sample. This study aligns with previous work involving the DOC that demonstrates a larger web presence of white individuals living in suburban or urban areas, which is not likely representative of the greater population with comorbid T1D and ASD [36]. This may represent a health disparity whereby those in rural areas or from diverse ethnic backgrounds may not be aware of the DOC, may not have access, or may not feel welcome or included in that space.

Conclusions

Raising a child with both ASD and T1D presents significant self-management challenges. However, caregivers of this population identify various coping techniques and strategies and find support in social media sites specific to this unique population. Further research is needed to develop new ways for the health care community to partner with this population.

Acknowledgments

This research was funded by the Penn State College of Medicine Qualitative Research Initiative. The authors would like to thank the caregivers of children with T1D and ASD for their participation in this study. The authors would also like to thank Dr Cheryl Tierney for reviewing the interview guide and Jessica Wright for contributing to study start-up and project management.

Conflicts of Interest

None declared.

References

1. Hilliard ME, Sparling KM, Hitchcock J, Oser TK, Hood KK. The emerging diabetes online community. Curr Diabetes Rev 2015;11(4):261-272 [FREE Full text] [doi: 10.2174/1573399811666150421123448] [Medline: 25901500]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Washington, DC, USA: American Psychiatric Association; 2013.
3. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveill Summ 2018 Apr 27;67(6):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6706a1] [Medline: 29701730]
4. Centers for Disease Control and Prevention. 2014. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States URL: https://stacks.cdc.gov/view/cdc/23442/cdc_23442_DS1.pdf? [accessed 2020-04-21]
5. Imperatore G, Boyle JP, Thompson TJ, Case D, Dabelea D, Hamman RF, SEARCH for Diabetes in Youth Study Group. Projections of type 1 and type 2 diabetes burden in the US population aged <20 years through 2050: dynamic modeling of incidence, mortality, and population growth. Diabetes Care 2012 Dec;35(12):2515-2520 [FREE Full text] [doi: 10.2337/dc12-0669] [Medline: 23173134]
6. Dabelea D, Mayer-Davis EJ, Saydah S, Imperatore G, Linder B, Divers J, SEARCH for Diabetes in Youth Study. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. J Am Med Assoc 2014 May 7;311(17):1778-1786 [FREE Full text] [doi: 10.1001/jama.2014.3201] [Medline: 24794371]
7. Centers for Disease Control and Prevention. 2019. Data & Statistics on Autism Spectrum Disorder URL: http://www.cdc.gov/ncbddd/autism/data.html [accessed 2015-06-30]
8. Xu G, Strathearn L, Liu B, Bao W. Prevalence of autism spectrum disorder among US children and adolescents, 2014-2016. J Am Med Assoc 2018 Jan 2;319(1):81-82 [FREE Full text] [doi: 10.1001/jama.2017.17812] [Medline: 29297068]
9. Stanek KR, Youngkin EM, Pyle LL, Raymond JK, Driscoll KA, Majidi S. Prevalence, characteristics, and diabetes management in children with comorbid autism spectrum disorder and type 1 diabetes. Pediatr Diabetes 2019 Aug;20(5):645-651. [doi: 10.1111/pedi.12848] [Medline: 30912248]
10. Bethin KE, Kanapka LG, Laffel LM, Majidi S, Chaytor NS, MacLeish S, T1D Exchange Clinic Network. Autism spectrum disorder in children with type 1 diabetes. Diabet Med 2019 Oct;36(10):1282-1286. [doi: 10.1111/dme.14069] [Medline: 31271682]
11. Taggart L, Coates V, Truesdale-Kennedy M. Management and quality indicators of diabetes mellitus in people with intellectual disabilities. J Intellect Disabil Res 2013 Dec;57(12):1152-1163. [doi: 10.1111/j.1365-2788.2012.01633.x] [Medline: 23106688]
12. Anwar AJ, Walker JD, Frier BM. Type 1 diabetes mellitus and Down's syndrome: prevalence, management and diabetic complications. Diabet Med 1998 Feb;15(2):160-163. [doi: 10.1002/(SICI)1096-9136(199802)15:2<160::AID-DIA537>3.0.CO;2-J] [Medline: 9507919]
13. Shireman TI, Reichard A, Nazir N, Backes JM, Greiner KA. Quality of diabetes care for adults with developmental disabilities. Disabil Health J 2010 Jul;3(3):179-185. [doi: 10.1016/j.dhjo.2009.10.004] [Medline: 21122783]
14. Janicki MP, Dalton AJ, Henderson CM, Davidson PW. Mortality and morbidity among older adults with intellectual disability: health services considerations. Disabil Rehabil 1999;21(5-6):284-294. [doi: 10.1080/096382899297710] [Medline: 10381241]
15. Rey-Conde T, Lennox N. Delivering diabetes care to people with intellectual disability. Diabetes Voice 2007;52(2):16-18 [FREE Full text]
16. Glanz K, Rimer BK, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, New Jersey: John Wiley & Sons; 2008.
17. Berry JG, Bloom S, Foley S, Palfrey JS. Health inequity in children and youth with chronic health conditions. Pediatrics 2010 Dec;126(Suppl 3):S111-S119. [doi: 10.1542/peds.2010-1466D] [Medline: 21123473]

18. Oser TK, Minnehan KA, Wong G, Parascando J, McGinley E, Radico J, et al. Using social media to broaden understanding of the barriers and facilitators to exercise in adults with type 1 diabetes. J Diabetes Sci Technol 2019 May;13(3):457-465 [FREE Full text] [doi: 10.1177/1932296819835787] [Medline: 30862185]
19. Oser TK, Oser SM, McGinley EL, Stuckey HL. A novel approach to identifying barriers and facilitators in raising a child with type 1 diabetes: qualitative analysis of caregiver blogs. JMIR Diabetes 2017 Oct 26;2(2):e27 [FREE Full text] [doi: 10.2196/diabetes.8966] [Medline: 30291073]
20. Sandelowski M. Using qualitative research. Qual Health Res 2004 Dec;14(10):1366-1386. [doi: 10.1177/1049732304269672] [Medline: 15538005]
21. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Jul 2;20(1):37-46. [doi: 10.1177/001316446002000104]
22. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 2;16(1):-. [doi: 10.1177/1609406917733847]
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
24. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, California: Sage Publications; 2017.
25. Harrison P, Oakland T. Pearson Assessments. 2015. Adaptive Behavior Assessment System URL: https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Behavior/Brief/Adaptive-Behavior-Assessment-System-%7C-Third-Edition/p/100001262.html [accessed 2020-04-21]
26. Harrison PL, Oakland T. Adaptive behavior assessment system. In: Kreutzer J, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, USA: Springer; 2017:1-4.
27. American Diabetes Association. Children and adolescents: standards of medical care in diabetes-2019. Diabetes Care 2019 Jan;42(Suppl 1):S148-S164. [doi: 10.2337/dc19-S013] [Medline: 30559239]
28. Berkman L, Glass T. Social integration, social networks, social support, and health. Soc Epidemiol 2000;1:137-173 [FREE Full text] [doi: 10.4135/9781412952576.n192]
29. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Soc Sci Med 2000 Sep;51(6):843-857. [doi: 10.1016/s0277-9536(00)00065-4] [Medline: 10972429]
30. Clayton KM, Stewart SM, Wiebe DJ, McConnel CE, Hughes CW, White PC. Maternal depressive symptoms predict adolescent healthcare utilization and charges in youth with type 1 diabetes (T1D). Health Psychol 2013 Sep;32(9):1013-1022. [doi: 10.1037/a0032761] [Medline: 24001252]
31. Dale JR, Williams SM, Bowyer V. What is the effect of peer support on diabetes outcomes in adults? A systematic review. Diabet Med 2012 Nov;29(11):1361-1377. [doi: 10.1111/j.1464-5491.2012.03749.x] [Medline: 22804713]
32. Eckshtain D, Ellis DA, Kolmodin K, Naar-King S. The effects of parental depression and parenting practices on depressive symptoms and metabolic control in urban youth with insulin dependent diabetes. J Pediatr Psychol 2010 May;35(4):426-435 [FREE Full text] [doi: 10.1093/jpepsy/jsp068] [Medline: 19710249]
33. Cunningham NR, Vesco AT, Dolan LM, Hood KK. From caregiver psychological distress to adolescent glycemic control: the mediating role of perceived burden around diabetes management. J Pediatr Psychol 2011 Mar;36(2):196-205 [FREE Full text] [doi: 10.1093/jpepsy/jsq071] [Medline: 20736389]
34. Maas-van Schaaijk NM, Roeleveld-Versteegh AB, van Baar AL. The interrelationships among paternal and maternal parenting stress, metabolic control, and depressive symptoms in adolescents with type 1 diabetes mellitus. J Pediatr Psychol 2013;38(1):30-40. [doi: 10.1093/jpepsy/jss096] [Medline: 22988060]
35. Litchman ML, Allen NA, McAdam-Marx C, Feehan M. Using projective exercises to identify patient perspectives of living with comorbid type 2 diabetes and asthma. Chronic Illn 2019 Sep 7:- epub ahead of print. [doi: 10.1177/1742395319872788] [Medline: 31495200]
36. Litchman ML, Walker HR, Ng AH, Wawrzynski SE, Oser SM, Greenwood DA, et al. State of the science: a scoping review and gap analysis of diabetes online communities. J Diabetes Sci Technol 2019 May;13(3):466-492 [FREE Full text] [doi: 10.1177/1932296819831042] [Medline: 30854884]

Abbreviations

ABAS-3: adaptive behavior assessment system, third edition
ASD: autism spectrum disorder
CGM: continuous glucose monitoring
DOC: diabetes online community
GAC: general adaptive composite
HbA_{1c}: hemoglobin A_{1c}
HIPAA: Health Insurance Portability and Accountability Act
ID: intellectual disability

REDCap: Research Electronic Data Capture
T1D: type 1 diabetes

Edited by G Eysenbach; submitted 24.11.19; peer-reviewed by A Ng, M Hilliard; comments to author 27.01.20; revised version received 22.03.20; accepted 26.03.20; published 03.06.20.

Please cite as:
Oser TK, Oser SM, Parascando JA, Grisolano LA, Krishna KB, Hale DE, Litchman M, Majidi S, Haidet P
Challenges and Successes in Raising a Child With Type 1 Diabetes and Autism Spectrum Disorder: Mixed Methods Study
J Med Internet Res 2020;22(6):e17184
URL: https://www.jmir.org/2020/6/e17184
doi:10.2196/17184
PMID:32217508

Review

Behavior Change Techniques Included in Reports of Social Media Interventions for Promoting Health Behaviors in Adults: Content Analysis Within a Systematic Review

Rosiane Simeon[1], MD, MPH; Omar Dewidar[1,2], HBsc; Jessica Trawin[1], BSocSci; Stephanie Duench[3], MPH; Heather Manson[4], MD, PhD; Jordi Pardo Pardo[5], BCom; Jennifer Petkovic[1], PhD; Janet Hatcher Roberts[1,2], MSc; Peter Tugwell[1,2,6,7], MD, PhD; Manosila Yoganathan[1], BA, MA; Justin Presseau[2,7,8], PhD; Vivian Welch[1,2], PhD

[1]Bruyère Research Institute, University of Ottawa, Ottawa, ON, Canada
[2]Faculty of Medicine, School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada
[3]Department of Population Medicine, University of Guelph, Guelph, ON, Canada
[4]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[5]Cochrane Musculoskeletal Group, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[6]Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[7]Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa, ON, Canada
[8]School of Psychology, Faculty of Social Sciences, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Rosiane Simeon, MD, MPH
Bruyère Research Institute
University of Ottawa
85 Primrose Ave
Ottawa, ON
Canada
Phone: 1 6138181154
Email: rsime006@uottawa.ca

Abstract

Background: Social media are an increasingly commonly used platform for delivering health promotion interventions. Although recent research has focused on the effectiveness of social media interventions for health promotion, very little is known about the optimal content within such interventions, and the active ingredients to promote health behavior change using social media are not clear. Identifying which behavior change techniques (BCTs) are reported may help to clarify the content of interventions using a generalizable terminology that may facilitate future intervention development.

Objective: This study aimed to identify which BCTs are reported in social media interventions for promoting health behavior change in adults.

Methods: We included 71 studies conducted with adult participants (aged ≥18 years) and for which social media intervention was considered interactive in a Cochrane review of the effectiveness of such interventions. We developed a coding manual informed by the Behavior Change Technique Taxonomy version 1 (BCTTv1) to identify BCTs in the included studies. We identified BCTs in all study arms (including control) and described BCTs in the group and self-directed components of studies. We characterized the dose of delivery for each BCT by low and high intensity. We used descriptive analyses to characterize the reported BCTs.

Results: Our data consisted of 71 studies published from 2001 to 2017, mainly conducted in high-income countries (n=65). Most studies (n=31) used tailored, interactive websites to deliver the intervention; Facebook was the most used mainstream platform. In developing our coding manual, we adapted some BCTTv1 instructions to better capture unique nuances of how BCTs were operationalized in social media with respect to likes, retweets, smiles, congratulations, and badges. Social support (unspecified), instruction on how to perform the behavior, and credible source were most frequently identified BCTs in intervention arms of studies and group-delivery settings, whereas instruction on how to perform the behavior was most commonly applied in self-directed components of studies, control arms, and individual participant settings. Instruction on how to perform the behavior was also the most frequently reported BCT in both intervention and control arms simultaneously. Instruction on how to perform the behavior, social support (unspecified), self-monitoring of behavior, information about health consequences, and credible source were identified in the top 5 BCTs delivered with the highest intensity.

Conclusions: This study within a review provides a detailed description of the BCTs and their dose to promote behavior change in web-based, interactive social media interventions. Clarifying active ingredients in social media interventions and the intensity of their delivery may help to develop future interventions that can more clearly build upon the existing evidence.

(J Med Internet Res 2020;22(6):e16002) doi:10.2196/16002

KEYWORDS

health behavior; taxonomy; social media; health promotion; public health

Introduction

Background

With more than 4 billion users worldwide and a projected yearly growth of 7% [1], the internet has become a central means of delivering health promotion interventions [2,3]. In particular, the advent and widespread use of social media have fostered a novel setting in which to deliver public health and health equity–oriented interventions [2,3]. These interventions greatly differ in their health aims, including HIV testing [4], mental health [5], physical activity [6], smoking cessation [7], and vaccination [8]. Social media can be characterized as consisting of “activities, practices, and behaviors among communities of people who gather online to share information, knowledge, and opinions using conversational media...that make it possible to create and easily transmit content in the form of words, pictures, videos, and audios” [9]. A primary goal of these social media interactions in health promotion is to change health behavior.

Interactions among participants in web-enabled interventions can lead to better health outcomes [10,11]. The past decade of research suggests that social media interventions may be effective in promoting health behavior change [12,13]. Earlier research has highlighted the need to improve the assessment of the design, implementation, and effectiveness of health promotion interventions delivered through social network technologies [14]. The findings from a recent review have revealed the diversity of social media features used in health-focused interventions delivered through the internet [15]. These features support several functions, including identity representation, peer grouping, and web-based social networking [15]. However, it is currently unclear which specific techniques are delivered within social media interventions, that is, which *active ingredients* are leveraged to promote behavior change in this new medium [11,12]. The active ingredient of an intervention that aims to change behavior can be defined as a behavior change technique (BCT), which consists of “an observable, replicable, and irreducible component of an intervention designed to alter or redirect causal processes that regulate behaviour” [16]. It is, therefore, important to describe the content of interventions delivered in social media and unpack their *active ingredients* to determine (1) what is actually being delivered in both intervention and control groups, (2) gaps and opportunities to consider integrating underrepresented BCTs, and (3) if this delivery mode of social media presents new BCTs or novel ways of operationalizing BCTs. Using a shared language to describe the intervention content, including which BCTs are used and how they are employed within an intervention, can help elucidate the development of future effective interventions.

Identification of BCTs using an agreed taxonomy ensures consistent comparison of techniques across studies, building toward a cumulative evidence base for informing terminology and for chartering the active ingredients of social media interventions. The Behavior Change Technique Taxonomy version 1 (BCTTv1) was developed to identify such active ingredients and to characterize the components of strategies that aim to change behavior [16-18]. The BCTTv1 with its 93 individual BCTs grouped into 16 categories [16] provides a standardized approach to classifying the content of health promotion interventions that involve behavior change. The taxonomy has been used extensively to clarify the active ingredients of health behavior change interventions in systematic reviews across several settings [19-21]. In addition, the effectiveness of health behavior change interventions in other areas has been shown to be associated with the inclusion of particular BCTs [22-24]. The findings from a systematic review that assessed the behavioral mechanisms of social media interventions on adolescent diet suggested that BCTs such as goal setting and self-monitoring of behavior may have a positive impact on changing dietary habits in adolescents [25]. However, the authors highlight the need to improve the description of BCTs delivered in social media interventions [25].

Objectives

In this study within a review (SWAR) [26], we aimed to gain a better understanding of the active ingredients of social media interventions by identifying and describing BCTs and assessing the applicability of the BCT framework to studies considered for a Cochrane review of the effectiveness of interactive social media interventions [27].

Methods

Identification of Studies

The details of the systematic review methods, including the search strategy, are available in our published protocol [27]. We searched the following major electronic databases: Cochrane Central Register of Controlled Trials (CENTRAL), Medical Literature Analysis and Retrieval System Online, or MEDLARS Online (MEDLINE), EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and PsycINFO. We also conducted a focused search for unpublished studies or reports within Google Scholar and Web of Science. The search was also extended to websites of public health governmental and nongovernmental organizations, such as the Public Health Agency of Canada, the World Health Organization (WHO), and international development agencies such as the Asian Development Bank and the Inter‐American Development Bank. Clinical trial registries and the WHO International Clinical Trials Registry Platform were also searched for relevant studies. We also searched the reference lists of the included studies. Multimedia Appendix 1 provides a sample of the search strategy and results from MEDLINE. In brief, we included studies focused on individuals from the general population who were aged ≥18 years with no limitations on their health status. We only included social media interventions that allowed two-way communication and interaction between and among participants. Some examples consisted of interventions supported through major (as of 2019) social media outlets, such as Facebook, Instagram, Twitter, and YouTube. However, interventions implemented through these platforms had to allow peer interaction to be considered in this review.

Study comparators included usual care, no intervention, or an active comparison (eg, one type of social media compared with another). The primary outcomes included validated measures of health-related behaviors, physical health, well-being, and psychological health. Studies were not excluded based on the outcome. The literature search was limited to studies that were published between January 2001 and March 2017 as most of the commonly used social media platforms were developed in 2001 or later (eg, Facebook and Twitter), and our overview showed no earlier studies using these (currently) widely used social media platforms [3]. The types of included studies consisted of randomized controlled trials (RCTs), controlled before-and-after studies, interrupted time series, and RCTs with stepped-wedge designs.

We excluded interventions that used one-way communication or one-to-one communication between a user and practitioner. For example, we excluded studies that assessed mobile health apps (eg, apps that track clinical information with contact between an individual and their health care provider) and studies with content that is transmitted unidirectionally (eg, text message reminder interventions in which the recipient is unable to reply and podcasts in which health information is provided with no opportunity for two-way communication) or which only allows for comments without sharing functionality, such as blogs. We also excluded studies that assessed web-based interventions that are based on the exchange between a single care provider and an individual participant, such as web-based cognitive behavioral therapy. Advertisements on social media (eg, Facebook) without interactive functionality and virtual gaming interventions were also excluded.

Coding and Analysis Strategies

For each eligible study, 2 trained coders (JT and RS) used BCTTv1 as a basis for coding BCTs. Coders completed BCTTv1 training and met the requirements for accurate and reliable coding. Following training, coders calibrated their coding by independently coding 3 included studies to resolve any initial discrepancies and to inform the development of a contextualized coding manual to supplement the BCTTv1. This was an iterative process to adapt the instructions if needed, which were applied to previously coded and future studies for testing. We developed a coding manual and identified any BCTs that were unique to social media. Multimedia Appendix 2 provides more details on the coding and analysis strategies related to this review.

We also distinguished between low- and high-intensity use to allow a balanced characterization of BCTs in interventions. In this review, we defined low-intensity BCT delivery as a given BCT delivered using one modality at one time. We coded high-intensity BCT delivery as a given BCT delivered more than once using the same or different modalities. Multimedia Appendix 2 provides examples of BCT statements coded as high or low intensity.

The titles of the included studies were entered in an Excel sheet, which contained all 93 BCTs organized under their respective categories. Our coding process captured the distribution of BCTs in interactive social media and self-directed components of the included studies. Interactive social media were defined as two-way communication and interaction between and among participants. Self-directed activities consisted of those carried out by participants outside of social media, such as tracking steps in a diary or using a pedometer. BCTs were extracted in both intervention and control arms in reported papers, appendices, websites, and other references indicated by the authors. We also recorded BCTs applied to participants in a group and individual participants. An interrater reliability score of kappa value 0.60 was used before starting to code [25]. Reviewers convened weekly to reconcile coding discrepancies. A third researcher (J Presseau) was consulted when discrepancies could not be resolved. Frequency counts were performed separately to quantify the categories of BCTs recorded and the intensity with which these were applied. A descriptive analysis was conducted on the identification and application of BCTs in interactive social media interventions.

Results

Characteristics of the Included Studies

Studies were identified using the search strategy specified in the systematic review protocol [27]. Our team screened more than 25,000 records and assessed 282 full-text studies for eligibility criteria. After applying the eligibility criteria, 71 studies were retained eligible for analysis. The majority of the studies consisted of RCTs (60 studies). As shown in Table 1, the studies reviewed were mainly conducted in high-income countries, which hosted 92% (65/71) of the interventions. Only 5 interventions were conducted in upper–middle-income countries. There was only 1 study identified in lower–middle-income countries.

Table 1. Summary of characteristics of included studies (N=71).

Study characteristics	Value, n (%)
Country	
High-income countries	65 (92)
Upper–middle-income countries	5 (7)
Lower–middle-income countries	1 (1)
Study design	
RCT[a]	60 (85)
Non-RCT	11 (15)
Types of social media	
Tailored interactive website only	31 (44)
Facebook only	23 (32)
Tailored interactive website and other social media components (Facebook and interactive apps)	13 (18)
Twitter or WhatsApp or WeChat	3 (4)
Twitter and other social media components	1 (1)
Types of outcomes[b]	
Health-related behaviors	40 (56)
Physical health	25 (35)
Well-being	11 (15)
Psychological health	8 (11)

[a] RCT: randomized controlled trial.

[b] More than one outcome per study.

Interactive social media attributes were identified in both the experimental and control arms of the included studies. In studies with active comparators, study arms were distinguished by changing the intensity of BCTs delivered [28,29], varying web-based social interactivity [6,30], or applying different BCTs in each arm [31,32]. The authors used a variety of interactive platforms to deliver interventions. The types of social media reported were distributed as follows: tailored, interactive websites (31 studies); Facebook only (23 studies); and a combination of tailored websites and other social media components (13 studies). Although not the focus of this review, self-directed strategies used in the intervention arms of the studies were noted. These included sharing informational and educational materials through means such as email [33-35] and video recordings [36,37]. Participants also had access to noninteractive aids such as pedometers [38,39] and used text messages [40,41] and personal diaries [42] to support their offline activities.

The included studies cover a variety of population groups, including college students [43], men who have sex with men [44], parents [45], patients with chronic diseases [46,47], and pregnant women [42,48].

We classified the outcomes of interest for the systematic review into 4 main categories: (1) health-related behaviors, (2) physical health, (3) well-being, and (4) psychological health. Health-related behaviors and physical health were the most reported outcomes and were identified in 40 and 25 studies, respectively. Health-related behaviors referred to the activities or practices taken by an individual, which either enhanced or deteriorated their health. It included measures such as physical activity [6,30,32,33,38,39,41,43,46,47,49-67], diet/nutrition [47,49,52,54,56,58,59,66,68,69], screening/testing [35,44,70-72], smoking [7,73-77], and medication adherence [45,78,79]. Physical health was used to describe any physiological measure such as BMI [32,41,56,58,63,67] and weight [38,40,48-50,52,59,65-67,69,80-84] but not biochemical markers of physiological health.

Well-being and psychological health were reported in 11 and 8 studies, respectively. Well-being was used to classify measures of quality of life [46,50,57,59,61,68,85], whereas psychological health was used to classify measures of psychological functioning, including depression [47,49,57,86,87], stress/distress [49,68,88], and anxiety [49,86].

Capturing Behavior Change Techniques Unique to Social Media

As shown in the examples in Table 2, in identifying BCTs across studies, 3 nuanced operationalizations of existing BCTs emerged that did not fit neatly within the existing taxonomy specifications without losing content. Our new instructions allowed us to capture overt endorsement and virtual rewards noted in specific interventions [67,89]. Overt endorsement consists of open approval of a message by participants, and it can be observed in behaviors such as forwarding emails and comments, likes, and *retweets* on Facebook and Twitter [47,48]. Virtual rewards consist of a system that allows collecting prizes in the form of *smiles*, *congratulations*, *badges*, *virtual gifts*, *or stars* to encourage participants in their progress toward achieving their goals [47,48].

Virtual rewards (by design), although in principle consistent with the *rewards (behavior)* BCT, have a level of nuance within the web-based environment that we decided it was worth making an explicit distinction in coding instructions for that BCT. Namely, we coded all statements that described overt endorsement such as *likes*, *smiles*, *virtual gifts*, *Twitter posts*, and *public post on profile page wall* as *social support (unspecified)* when these statements came from participants themselves in a natural manner. However, we coded these same statements as *nonspecific reward* or *social reward* when they were embedded in the design of the intervention to acknowledge the achievement of participants.

Table 2. Application of newly created rules for capturing virtual rewards.

New instructions	Representative example of coded statements	Behavior change technique taxonomy version 1 code
Overt endorsement (from participants)	"Participants in this intervention were free to post their own relevant messages through videos, text, pictures and so on. They could engage in relevant discussions via commenting." [90]	Social support (unspecified)
Virtual rewards (by design)	"It included gamification features, such as awards for individual and team step-logging and step-count achievements and the ability to send virtual gifts to teammates." [61]	Social reward
Virtual rewards (by design)	"Engagement was rewarded in the intervention group with points, badges and gradual revealing of graphic-level images and other virtual elements." [89]	Nonspecific reward

Identification of Behavior Change Techniques in the Studies

We identified 46 out of 93 techniques from the BCTTv1 across the 71 included studies, which are displayed in Multimedia Appendix 3. BCTs were mainly reported in the Methods section of the studies (68 studies). We also identified BCTs from additional sources, such as protocols (11 studies), study appendices (4 studies), and author-identified companion papers (4 studies).

Behavior Change Techniques Applied in Interactive Social Media and Self-Directed Components

More than 55% (26/46) of the BCTs identified were applied in the interactive social media component of interventions (two-way communication and interaction between and among participants). The top 15 BCTs applied across interactive social media and self-directed components of studies (one-way communication or no social interaction) are displayed in Table 3. As expected, *social support (unspecified)* was the most common BCT identified in interactive social media components (51 studies) [6-8,28,30,32,35,36,38,39,41,44-47,50-55,57,59,61, 63-66,69-73,75,77-80,82,83,85,87-95], followed by *instruction onhow to perform the behavior* (21 studies) [28,36,37, 46-49,52,61,64,68,72,76,77,82,84,85,91,95-97], *credible source* (16 studies) [7,28,30,35,36,39,47,55,64,69,72,76,79,94,95,97], *social comparison* (8 studies) [36,43,48,54,61,65,80,82], and *information about health consequences* (8 studies) [28,43,64,68,72,83,90,95]. On the other hand, *instruction on how to perform the behavior* was the most commonly identified BCT in the self-directed component of interventions (39 studies) [6-8,30,32,34,38-41,43-45,50,51,53,55-60,62,65-67,69,71,74,75,79-81, 83,86,87,89,93,94], followed by *self-monitoring of behavior* (32 studies) [6,30,32,33,38-41,45,46,48,50,51,54-58,60,63-66, 69,76,79,81-84,92,98], *credible source* (29 studies) [8,32,34,38,41,44,45,48,50,53,56,58,59,62,65,66,68, 74,75,77,80-82,85,88,90,92,93,96], *goal-setting behavior* (25 studies) [6,30,32,38,39,41,48,50,51,54,55,57-60,65-67,74,75, 81-84,98], and *adding objects to the environment* (24 studies) [30,32,38-41,46,48-51,56-61,66,70,80,81,83,93,98].

Table 3. Top 15 behavior change techniques captured in the components of interventions (n=71).

Behavior change techniques	Social media, n[a] (%)	Self-directed, n[a] (%)
1.1 Goal setting (behavior)	7 (10)	25 (35)
1.2 Problem solving	6 (8)	13 (18)
1.4 Action planning	N/A[b]	19 (27)
2.2 Feedback on behavior	3 (4)	19 (27)
2.3 Self-monitoring of behavior	5 (7)	32 (45)
2.4 Self-monitoring of outcome(s) of behavior	2 (3)	N/A
3.1 Social support (unspecified)	51 (72)	12 (17)
3.2 Social support (practical)	7 (10)	17 (24)
3.3 Social support (emotional)	3 (4)	N/A
4.1 Instruction on how to perform the behavior	21 (30)	39 (55)
5.1 Information about health consequences	8 (11)	17 (24)
6.1 Demonstration of the behavior	7 (10)	11 (15)
6.2 Social comparison	8 (11)	N/A
7.1 Prompts/cues	7 (10)	21 (30)
8.7 Graded tasks	N/A	14 (20)
9.1 Credible source	16 (23)	29 (41)
11.1 Pharmacological support	N/A	9 (13)
12.2 Restructuring the social environment	5 (7)	N/A
12.5 Adding objects to the environment	N/A	24 (34)

[a] Behavior change techniques applied in interactive social media components and self-directed components.

[b] N/A: not applicable.

Behavior Change Techniques Applied in the Intervention and Control Arms

The top 5 BCTs applied in intervention arms of the included studies: *social support (unspecified*; 56 studies) [6-8,28,30,32-36,38-41,44-47,52-62,64-66,69,70,72,73,75,77-83,85-95,97], *instruction on how to perform the behavior* (37 studies) [8,28,36,37,40,44-48,52,53,56-58,61,64,67,68,72,75,77,79,81-87,89,91,93-97], *credible source* (35 studies) [7,8,28,34-36,39,41,44,45,47,48,53,56,58,59,62,64,65,68,69,72,75,77,79,81,82,85,88,92-97], *self-monitoring of behavior* (25 studies) [6,32,39,40,45-48,56-58,61,64,65,69,76,78-84,88,92], and *prompts/cues* (24 studies) [33,39-41,43,47,48,52,55,56,58,61,64,65,69,73,78,79,81,82,89,92,93,97]. The content of the control arms was generally reported as *usual care*. Only 9 different BCTs were uniquely applied in the control arms of the studies. *Instruction on how to perform the behavior*, *social support (unspecified)*, and *information about health consequences* constituted the most commonly applied BCTs in control arms only and were identified in 8 [7,29,34,38,39,41,59,66], 3 [50,51,63], and 2 [38,63] studies, respectively. The top 15 BCTs applied to the intervention and control arms of the included studies are displayed in Table 4.

Across the included studies, 14 of the top 15 BCTs applied in intervention arms only were also used in studies with BCTs applied in both arms. The exception was *social comparison*, which was solely identified in the intervention arms of 11 studies [32,33,36,43,48,54,58,61,65,80,82]. As shown in Table 4, the ranking of BCTs also changed with the simultaneous application of BCTs in both study arms in comparison with application in intervention arms only.

Table 4. Top 15 behavior change techniques in the study arms (n=71).

Behavior change techniques	Intervention, n[a] (%)	Control, n[a] (%)	Intervention and control, n[a] (%)
1.1 Goal setting (behavior)	19 (27)	N/A[b]	13 (18)
1.2 Problem solving	13 (18)	1 (1)	5 (7)
1.4 Action planning	14 (20)	1 (1)	6 (8)
2.2 Feedback on behavior	16 (23)	N/A	6 (8)
2.3 Self-monitoring of behavior	25 (35)	1 (1)	11 (15)
3.1 Social support (unspecified)	56 (79)	3 (4)	4 (6)
3.2 Social support (practical)	18 (25)	N/A	6 (8)
4.1 Instruction on how to perform the behavior	37 (52)	8 (11)	15 (21)
5.1 Information about health consequences	15 (21)	2 (3)	8 (11)
6.1 Demonstration of the behavior	14 (20)	N/A	4 (6)
6.2 Social comparison	11 (25)	N/A	N/A
7.1 Prompts/cues	21 (30)	N/A	7 (10)
8.7 Graded tasks	9 (13)	N/A	5 (7)
9.1 Credible source	35 (49)	1 (1)	9 (13)
11.1 Pharmacological support	N/A	N/A	4 (6)
11.3 Conserving mental resources	N/A	1 (1)	N/A
12.5 Adding objects to the environment	13 (18)	N/A	11 (15)
13.2 Framing/reframing	N/A	1 (1)	N/A

[a] Behavior change techniques applied in intervention arms only, control arms, and intervention and control arms simultaneously.

[b] N/A: not applicable.

Behavior Change Techniques Applied to Group and Individual Participants

Social support (unspecified) was the most commonly applied BCTs to a group of participants (50 studies) [6-8,30,32,33,36,38,39,41,44-46,50-61,63-66,69-73,75-78,80,82,83,85-87,90-95]. It was then followed by *instruction on how to perform the behavior* (24 studies) [36,37,44,45,48,52,53,55-58,62,64,67,72,76,82,84,85,87,93,95-97], *credible source* (23 studies) [7,30,35,36,39,44,45,47,53,55,56,58,64,69,72,75,77,79,85,90,94,95,97], *demonstration of the behavior* (12 studies) [6,36,52,53,55,56,58,62,64,76,93,94], and *social support practical* (11 studies) [7,32,33,39,48,52,53,58,68,72,76]. When BCTs were applied to individuals, *instruction on how to perform the behavior* was the most frequent (36 studies) [6-8,28,30,32,34,38-41,43,46,47,49-51,59-61,65,66,68,69,71,74,75,77,79-81,83,86,89,91,94], followed by *self-monitoring of behavior* (34 studies) [6,30,32,33,38-41,46-48,50,51,54-58,60,61,63,65,66,69,76,78-83,88,92,98], *goal-setting behavior* (30 studies) [6,30,32,38,39,41,46-48,50,51,54,55,57-60,65-67,74-77,81-84,86,98], *adding objects to the environment* (24 studies) [30,32,38-41,46,48-51,56-61,66,70,80,81,83,93,98], *and credible source* (22 studies) [8,28,32,34,38,41,48,50,59,62,65,66,68,74,77,80-82,88,92,93,96]. Table 5 shows the most frequent BCTs applied to group and individual participants.

Table 5. Top 15 behavior change techniques in participant settings (n=71).

Behavior change techniques	Group setting, n[a] (%)	Individual setting, n[a] (%)
1.1 Goal setting (behavior)	N/A[b]	30 (42)
1.2 Problem solving	5 (7)	14 (20)
1.4 Action planning	5 (7)	16 (23)
1.5 Review behavior goal(s)	N/A	10 (14)
2.2 Feedback on behavior	N/A	19 (27)
2.3 Self-monitoring of behavior	3 (4)	34 (48)
2.8 Feedback on outcome(s) of behavior	N/A	7 (10)
3.1 Social support (unspecified)	50 (70)	13 (18)
3.2 Social support (practical)	11 (15)	13 (18)
3.3 Social support (emotional)	3 (4)	N/A
4.1 Instruction on how to perform the behavior	24 (34)	36 (51)
5.1 Information about health consequences	6 (8)	19 (27)
5.3 Information about social and environmental consequences	3 (4)	N/A
6.1 Demonstration of the behavior	12 (17)	N/A
6.2 Social comparison	9 (13)	N/A
7.1 Prompts/cues	9 (13)	19 (27)
8.1 Behavioral practice/rehearsal	4 (6)	N/A
8.7 Graded tasks	N/A	13 (18)
9.1 Credible source	23 (32)	22 (31)
12.2 Restructuring the social environment	6 (8)	N/A
12.5 Adding objects to the environment	N/A	24 (34)

[a] Behavior change techniques applied to individuals in the group setting and individual participants.

Low- and High-Intensity Use of Behavior Change Techniques

We characterized the intensity of delivering BCTs by identifying the modalities used to expose participants to a BCT. BCTs that were delivered using at least two modalities were coded as high intensity. Participants were exposed to BCTs through interactive social media and self-directed components in intervention or control arms of the studies and in a group or individually. The top 5 BCTs that were delivered with high intensity included *instruction on how to perform the behavior* (35 studies) [7,28,32,34,36-38,40,45,48,49,52,53,55-59,62,65-69,72,74,75,77,79,80,83,85,93,95,96], social *support (unspecified*; 28 studies) [7,8,28,32-35,37,40,41,45,50,51,56,61-67,71,73-75,77,81,86], *self-monitoring of behavior* (16 studies) [32,50,51,54,57,58,60,63,64,66,67,69,78,82,92,96], *information about health consequences* (13 studies) [7,32,34,36,38,48,50,56,59,80,81,95,96], and *credible source* (13 studies) [7,32,34,36,38,48,50,56,59,80,81,95,96]. The top 15 most frequent BCTs coded for high-intensity delivery are displayed in Table 6.

Table 6. Top 15 behavior change techniques coded for high-intensity delivery (n=71).

Behavior change techniques	Value, n (%)
1.1 Goal setting (behavior)	10 (14)
1.2 Problem solving	5 (7)
1.4 Action planning	8 (11)
1.5 Review behavior goal(s)	3 (4)
12.5 Adding objects to the environment	8 (11)
2.2 Feedback on behavior	8 (11)
2.3 Self-monitoring of behavior	16 (23)
3.1 Social support (unspecified)	28 (39)
3.2 Social support (practical)	4 (6)
4.1 Instruction on how to perform the behavior	35 (49)
5.1 Information about health consequences	15 (21)
6.1 Demonstration of the behavior	6 (8)
6.2 Social comparison	3 (4)
7.1 Prompts/cues	3 (4)
9.1 Credible source	13 (18)

Discussion

Principal Findings

We found that the BCTTv1 was applicable to studies promoting health behaviors using social media in a web-based environment. Given that all these studies focused on interactive web-based platforms, it is surprising that relatively few studies have reported leveraging the full spectrum of social comparison and social support BCTs. Furthermore, few studies have reported using BCTs that have been associated with effective behavior change in other settings (eg, goal setting and action planning) [99,100].

We made 3 adaptations to the coding structure to accurately and specifically code overt endorsement and virtual rewards, which we decided to be a type of social support when provided by peers through the internet and a type of social reward when provided by intervention designers. We also developed a method for coding the intensity of BCTs, which allowed us to identify studies with more intense use of BCTs. We added to the description of some BCTs to capture these nuances of interactive social media. Kadushin [101] and Southwell [102] highlighted the potential of overt endorsement and virtual rewards in influencing individual attitudes and health-related behaviors. We also noted that these techniques were reported in other studies as a measurement of engagement with participants and reach of interventions [67,90].

The interventions were mainly delivered through tailored, interactive platforms (ie, platforms developed by the research team). At the time of conducting this research, Facebook remained the most commonly used platform among the mainstream social media outlets for intervention studies and was used either alone or in combination with other social media components. Nonetheless, this does not dismiss the popularity of other social media platforms implemented in other countries. Social media platforms usually allow social interactions beyond geographical boundaries. We noted that none of the included studies had a global focus, that is, all were conducted within a specific country or region. In addition, with the widespread digital technologies, we expected to see more than one study from low- and middle-income countries. Resource settings [103] and research capacity [104,105] might be explored when assessing possible reasons for limited implementation internationally.

Almost all the studies used a variety of BCTs to deliver interventions through social media. The identified BCTs were applied in both the intervention and control arms of the studies. For this reason, it is essential to identify the intensity with which BCTs are applied. Identifying the intensity can help to determine if the extent of exposure makes a difference in the impact of changing behavior. Even after considering intensity, caution should still be applied when assessing the role of BCTs in changing the behavior of participants because some BCTs were applied with almost equal intensity in the control and intervention arms. This was observed for the BCT *instruction on how to perform the behavior.*

In addition, the content of control arms was mainly reported as *usual care*, with often no accompanying details on the features of such conditions. This situation makes it difficult to adequately characterize the content of control arms and warrants more analysis in attributing effectiveness to BCTs applied in intervention arms [52]. The types of BCTs applied to individuals in a group setting and those applied to individual participants varied greatly. This provides opportunities to explore which BCTs are optimal in a group setting and to identify which ones are inherently appropriate for individual activity only. Additional research can help elucidate the mechanisms of action of BCTs in interventions and further contribute to their optimal use to foster behavior change.

Limitations

One limitation of this research is that we relied on what was reported in the source studies. We attempted to check the web-based sites, but a lot had changed since the publication of the studies in the review. Therefore, the relative paucity of some BCTs (eg, social comparison and goal setting) may be related to insufficient detail in the published papers to code these aspects of the intervention. We encourage future reports of interventions to provide screenshots or access to the version of the platform used at the time of the intervention as additional files or links. In addition, the use of validated tools such as the Consolidated Standards of Reporting Trials eHealth can help improve the description and reporting of web-based interventions, thereby facilitating their implementation and replication to improve health outcomes [106].

Future research can also contribute to better characterize BCTs applied to foster social support in interactive social media interventions. *Social support (unspecified)* was the most commonly applied BCT in the studies analyzed. The definition of this BCT is comprehensive enough to allow the identification of social assistance with no specific descriptors, whereas BCTs such as *social support (practical)* and *social support (emotional)* allow capturing what social assistance entails. A better characterization of *social support* in interventions could help to distill in what ways this particular social interaction makes a difference in the behavior of participants.

Comparisons With Prior Work

A recent review that focused on social media features in web-based interventions reported that communication in the form of forums is the most common social media feature for sustaining behavior change [15]. These results are aligned with those of this SWAR, which shows that generic social support is the most common BCT used in interactive social media interventions, which was mainly captured in interventions that use discussion forums, group activities, and other exchanges to foster mutual support among participants.

Future Work

In this study, we provide an overview of the most commonly reported BCTs applied in social media intervention studies. As social media interventions continue to increase in popularity, it may be useful in future work to assess how the presence of particular BCTs or a combination of BCTs is associated with effects on health behavior change.

Conclusions

Our newly developed coding instructions along with BCTTv1 were useful in characterizing behavior change content in social media interventions and evaluating its effectiveness. This assessment of BCTs identified nuances in the operationalization of the BCTTv1 to characterize the unique features of interactive social media interventions for health promotion. These characteristics may be considered when designing social media platforms aimed at promoting public health. Our coding could further help deconstruct the complexity of interventions that use digital technologies and provide further understanding of the role of specific BCTs on behavior change. Although other studies highlighted the effectiveness of BCTs, such as goal setting and action planning, very few of the included studies used them. Increasing the use of BCTs is known to be effective in improving social media interventions.

Acknowledgments

The authors would like to thank Tamara Rader who developed the search strategy for the systematic review with the assistance of Jordi Pardo and Manosila Yoganathan. The authors acknowledge the assistance of Alexandra Sokolovski in testing the data extraction form. RS received funding from the Bruyère Research Institute Studentship Program. PT is the Canada Research Chair of Health Equity. VW is a recipient of the Ontario Early Research Award.

Authors' Contributions

RS, VW, J Presseau, JT, and OD drafted the review. Input on the review draft was provided by HM, SD, JR, J Petkovic, MY, J Presseau, and PT. VW and J Presseau are joint senior authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
User manual for coding behavior change techniques using the Behavior Change Technique Taxonomy version 1 in social media–based health interventions.
[DOCX File , 14 KB - jmir_v22i6e16002_app1.docx]

Multimedia Appendix 2
Sample of search strategy and results from the Medical Literature Analysis and Retrieval System Online, or MEDLARS Online database.
[DOCX File , 41 KB - jmir_v22i6e16002_app2.docx]

Multimedia Appendix 3
Distribution of behavior changing techniques captured across the 71 included studies.
[PNG File , 484 KB - jmir_v22i6e16002_app3.png]

References

1. We Are Social. 2018. Digital in 2018: World's Internet Users Pass the 4 Billion Mark URL: https://wearesocial.com/blog/2018/01/global-digital-report-2018 [accessed 2019-05-27]
2. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
3. Welch V, Petkovic J, Pardo JP, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
4. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
5. de Choudhury M, Kıcıman E. The language of social support in social media and its effect on suicidal ideation risk. Proc Int AAAI Conf Weblogs Soc Media 2017 May;2017:32-41 [FREE Full text] [Medline: 28840079]
6. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
7. Ramo DE, Thrul J, Chavez K, Delucchi KL, Prochaska JJ. Feasibility and quit rates of the tobacco status project: a Facebook smoking cessation intervention for young adults. J Med Internet Res 2015 Dec 31;17(12):e291 [FREE Full text] [doi: 10.2196/jmir.5209] [Medline: 26721211]
8. Glanz JM, Wagner NM, Narwaney KJ, Kraus CR, Shoup JA, Xu S, et al. Web-based social media intervention to increase vaccine acceptance: a randomized controlled trial. Pediatrics 2017 Dec;140(6):e20171117 [FREE Full text] [doi: 10.1542/peds.2017-1117] [Medline: 29109107]
9. Safko L. The Social Media Bible: Tactics, Tools, and Strategies for Business Success. Third Edition. Hoboken, New Jersey: Wiley; 2012.
10. Jones K, Baldwin KA, Lewis PR. The potential influence of a social media intervention on risky sexual behavior and Chlamydia incidence. J Community Health Nurs 2012;29(2):106-120. [doi: 10.1080/07370016.2012.670579] [Medline: 22536914]
11. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
12. Maher CA, Lewis LK, Ferrar K, Marshall S, de Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
13. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
14. Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. J Med Internet Res 2015 Jun 11;17(6):e141 [FREE Full text] [doi: 10.2196/jmir.3662] [Medline: 26068087]
15. Elaheebocus SM, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018 Feb 22;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
16. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
17. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
18. Nudelman G, Shiloh S. Mapping health behaviors: constructing and validating a common-sense taxonomy of health behaviors. Soc Sci Med 2015 Dec;146:1-10. [doi: 10.1016/j.socscimed.2015.10.004] [Medline: 26473449]
19. Presseau J, Ivers NM, Newham JJ, Knittle K, Danko KJ, Grimshaw JM. Using a behaviour change techniques taxonomy to identify active ingredients within trials of implementation interventions for diabetes care. Implement Sci 2015 Apr 23;10:55 [FREE Full text] [doi: 10.1186/s13012-015-0248-7] [Medline: 25900104]
20. Fulton EA, Brown KE, Kwah KL, Wild S. StopApp: using the behaviour change wheel to develop an app to increase uptake and attendance at NHS stop smoking services. Healthcare (Basel) 2016 Jun 8;4(2):e31 [FREE Full text] [doi: 10.3390/healthcare4020031] [Medline: 27417619]

21. Michie S, West R, Sheals K, Godinho CA. Evaluating the effectiveness of behavior change techniques in health-related behavior: a scoping review of methods used. Transl Behav Med 2018 Mar 1;8(2):212-224 [FREE Full text] [doi: 10.1093/tbm/ibx019] [Medline: 29381786]
22. Albarracín D, Gillette JC, Earl AN, Glasman LR, Durantini MR, Ho M. A test of major assumptions about behavior change: a comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. Psychol Bull 2005 Nov;131(6):856-897 [FREE Full text] [doi: 10.1037/0033-2909.131.6.856] [Medline: 16351327]
23. Fortune N, Madden R, Almborg A. Use of a new international classification of health interventions for capturing information on health interventions relevant to people with disabilities. Int J Environ Res Public Health 2018 Jan 17;15(1):145 [FREE Full text] [doi: 10.3390/ijerph15010145] [Medline: 29342077]
24. Rothman AJ. 'Is there nothing more practical than a good theory?': why innovations and advances in health behavior change will arise if interventions are used to test and refine theory. Int J Behav Nutr Phys Act 2004 Jul 27;1(1):11 [FREE Full text] [doi: 10.1186/1479-5868-1-11] [Medline: 15279674]
25. Hsu MS, Rouf A, Allman-Farinelli M. Effectiveness and behavioral mechanisms of social media interventions for positive nutrition behaviors in adolescents: a systematic review. J Adolesc Health 2018 Nov;63(5):531-545. [doi: 10.1016/j.jadohealth.2018.06.009] [Medline: 30197198]
26. -. Education section-studies within a review (SWAR). J Evid Based Med 2012 Aug;5(3):188-189. [doi: 10.1111/j.1756-5391.2012.01193.x] [Medline: 23672229]
27. Petkovic J, Simeon R, Presseau J, Gagnon D, Hossain A, Pardo PJ. Interactive social media interventions for health behaviour change, health outcomes, and health equity in the adult population. Cochrane Database Syst Rev 2018(2):CD012932. [doi: 10.1002/14651858.CD012932]
28. Mayer AB, Harrison JA. Safe Eats: an evaluation of the use of social media for food safety education. J Food Prot 2012 Aug;75(8):1453-1463. [doi: 10.4315/0362-028X.11-551] [Medline: 22856569]
29. Schaller A, Petrowski K, Pfoertner T, Froboese I. Effectiveness of a theory-based multicomponent intervention (movement coaching) on the promotion of total and domain-specific physical activity: a randomised controlled trial in low back pain patients. BMC Musculoskelet Disord 2017 Nov 6;18(1):431 [FREE Full text] [doi: 10.1186/s12891-017-1788-6] [Medline: 29110703]
30. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
31. Graham A. Development of the pharyngeal arches. Am J Med Genet A 2003 Jun 15;119A(3):251-256. [doi: 10.1002/ajmg.a.10980] [Medline: 12784288]
32. Rovniak LS, Kong L, Hovell MF, Ding D, Sallis JF, Ray CA, et al. Engineering online and in-person social networks for physical activity: a randomized trial. Ann Behav Med 2016 Dec;50(6):885-897 [FREE Full text] [doi: 10.1007/s12160-016-9814-8] [Medline: 27405724]
33. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: a randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
34. O'Neil PM, Miller-Kovach K, Tuerk PW, Becker LE, Wadden TA, Fujioka K, et al. Randomized controlled trial of a nationally available weight control program tailored for adults with type 2 diabetes. Obesity (Silver Spring) 2016 Nov;24(11):2269-2277 [FREE Full text] [doi: 10.1002/oby.21616] [Medline: 27804264]
35. Hwang KO, Ottenbacher AJ, Graham AL, Thomas EJ, Street RL, Vernon SW. Online narratives and peer support for colorectal cancer screening: a pilot randomized trial. Am J Prev Med 2013 Jul;45(1):98-107. [doi: 10.1016/j.amepre.2013.02.024] [Medline: 23790994]
36. Fiks AG, Gruver RS, Bishop-Gilyard CT, Shults J, Virudachalam S, Suh AW, et al. A social media peer group for mothers to prevent obesity from infancy: the Grow2Gether randomized trial. Child Obes 2017 Oct;13(5):356-368 [FREE Full text] [doi: 10.1089/chi.2017.0042] [Medline: 28557558]
37. Laws RA, Denney-Wilson EA, Taki S, Russell CG, Zheng M, Litterbach E, et al. Key lessons and impact of the growing healthy mhealth program on milk feeding, timing of introduction of solids, and infant growth: quasi-experimental study. JMIR Mhealth Uhealth 2018 Apr 19;6(4):e78 [FREE Full text] [doi: 10.2196/mhealth.9040] [Medline: 29674313]
38. Rote AE, Klos LA, Brondino MJ, Harley AE, Swartz AM. The efficacy of a walking intervention using social media to increase physical activity: a randomized trial. J Phys Act Health 2015 Jun 16;12(Suppl 1):S18-S25. [doi: 10.1123/jpah.2014-0279] [Medline: 25599378]
39. Schneider KL, Murphy D, Ferrara C, Oleski J, Panza E, Savage C, et al. An online social network to increase walking in dog owners: a randomized trial. Med Sci Sports Exerc 2015 Mar;47(3):631-639 [FREE Full text] [doi: 10.1249/MSS.0000000000000441] [Medline: 25003777]
40. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31 [FREE Full text] [doi: 10.1002/oby.20232] [Medline: 23505165]

41. Joseph RP, Keller C, Adams MA, Ainsworth BE. Print versus a culturally-relevant Facebook and text message delivered intervention to promote physical activity in African American women: a randomized pilot trial. BMC Womens Health 2015 Mar 27;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]
42. Linden K, Berg M, Adolfsson A, Sparud-Lundin C. Person-centred, web-based support in pregnancy and early motherhood for women with type 1 diabetes mellitus: a randomized controlled trial. Diabet Med 2018 Feb;35(2):232-241 [FREE Full text] [doi: 10.1111/dme.13552] [Medline: 29171071]
43. Al-Eisa E, Al-Rushud A, Alghadir A, Anwer S, Al-Harbi B, Al-Sughaier N, et al. Effect of motivation by 'Instagram' on adherence to physical activity among female college students. Biomed Res Int 2016;2016:1546013 [FREE Full text] [doi: 10.1155/2016/1546013] [Medline: 27034927]
44. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of internet popular opinion leaders (iPOL) among internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
45. Bai G, Wang Y, Yang L, Niu W. Effectiveness of a focused, brief psychoeducation program for parents of ADHD children: improvement of medication adherence and symptoms. Neuropsychiatr Dis Treat 2015;11:2721-2735 [FREE Full text] [doi: 10.2147/NDT.S88625] [Medline: 26604761]
46. Moy ML, Collins RJ, Martinez CH, Kadri R, Roman P, Holleman RG, et al. An internet-mediated pedometer-based program improves health-related quality-of-life domains and daily step counts in COPD: a randomized controlled trial. Chest 2015 Jul;148(1):128-137 [FREE Full text] [doi: 10.1378/chest.14-1466] [Medline: 25811395]
47. Bantum EO, Albright CL, White KK, Berenberg JL, Layi G, Ritter PL, et al. Surviving and thriving with cancer using a web-based health behavior change intervention: randomized controlled trial. J Med Internet Res 2014 Feb 24;16(2):e54 [FREE Full text] [doi: 10.2196/jmir.3020] [Medline: 24566820]
48. Herring SJ, Cruice JF, Bennett GG, Davey A, Foster GD. Using technology to promote postpartum weight loss in urban, low-income mothers: a pilot randomized controlled trial. J Nutr Educ Behav 2014;46(6):610-615 [FREE Full text] [doi: 10.1016/j.jneb.2014.06.002] [Medline: 25069621]
49. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
50. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013 Sep;7(3):355-368 [FREE Full text] [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]
51. Mailey E, Huberty J, Irwin B. Feasibility and effectiveness of a web-based physical activity intervention for working mothers. J Phys Act Health 2016 Aug;13(8):822-829. [doi: 10.1123/jpah.2015-0643] [Medline: 26999823]
52. Dadkhah M. Mountain Scholar Home. 2014. Development and Evaluation of the America on the Move Program For University Students URL: https://mountainscholar.org/bitstream/handle/10217/80937/Dadkhah_colostate_0053A_12013.pdf?sequence [accessed 2020-04-27]
53. John WC, Leng H, Kee Y. Use of Facebook in physical activity intervention programme: a test of self-determination theory. Int J Sport Psychol 2015;46(3):210-224. [doi: 10.7352/IJSP.2015.46.210]
54. Duncan M, Vandelanotte C, Kolt GS, Rosenkranz RR, Caperchione CM, George ES, et al. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-aged males: randomized controlled trial of the ManUp study. J Med Internet Res 2014 Jun 12;16(6):e136 [FREE Full text] [doi: 10.2196/jmir.3107] [Medline: 24927299]
55. Looyestyn J, Kernot J, Boshoff K, Maher C. A web-based, social networking beginners' running intervention for adults aged 18 to 50 years delivered via a Facebook group: randomized controlled trial. J Med Internet Res 2018 Feb 26;20(2):e67 [FREE Full text] [doi: 10.2196/jmir.7862] [Medline: 29483065]
56. Ling J, Robbins LB, Zhang N, Kerver JM, Lyons H, Wieber N, et al. Using Facebook in a healthy lifestyle intervention: feasibility and preliminary efficacy. West J Nurs Res 2018 Dec;40(12):1818-1842. [doi: 10.1177/0193945918756870] [Medline: 29421989]
57. Wan ES, Kantorowski A, Homsy D, Teylan M, Kadri R, Richardson CR, et al. Promoting physical activity in COPD: insights from a randomized trial of a web-based intervention and pedometer use. Respir Med 2017 Sep;130:102-110 [FREE Full text] [doi: 10.1016/j.rmed.2017.07.057] [Medline: 29206627]
58. Petrella RJ, Gill DP, Zou G, de Cruz A, Riggin B, Bartol C, et al. Hockey fans in training: a pilot pragmatic randomized controlled trial. Med Sci Sports Exerc 2017 Dec;49(12):2506-2516 [FREE Full text] [doi: 10.1249/MSS.0000000000001380] [Medline: 28719494]
59. Jane M, Hagger M, Foster J, Ho S, Kane R, Pal S. Effects of a weight management program delivered by social media on weight and metabolic syndrome risk factors in overweight and obese adults: a randomised controlled trial. PLoS One 2017;12(6):e0178326 [FREE Full text] [doi: 10.1371/journal.pone.0178326] [Medline: 28575048]
60. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]

61. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, de Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. J Med Internet Res 2015 Jul 13;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
62. Schaller A, Dintsios C, Icks A, Reibling N, Froboese I. Promoting physical activity in low back pain patients: six months follow-up of a randomised controlled trial comparing a multicomponent intervention with a low intensity intervention. Clin Rehabil 2016 Sep;30(9):865-877 [FREE Full text] [doi: 10.1177/0269215515618730] [Medline: 27496696]
63. Vandelanotte C, Kolt GS, Caperchione CM, Savage TN, Rosenkranz RR, Maeder AJ, et al. Effectiveness of a web 2.0 intervention to increase physical activity in real-world settings: randomized ecological trial. J Med Internet Res 2017 Nov 13;19(11):e390 [FREE Full text] [doi: 10.2196/jmir.8484] [Medline: 29133282]
64. Coffeng JK, Boot CR, Duijts SF, Twisk JW, van Mechelen W, Hendriksen IJ. Effectiveness of a worksite social & physical environment intervention on need for recovery, physical activity and relaxation; results of a randomized controlled trial. PLoS One 2014;9(12):e114860 [FREE Full text] [doi: 10.1371/journal.pone.0114860] [Medline: 25542039]
65. Willcox J, Wilkinson S, Lappas M, Ball K, Crawford D, McCarthy E, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728 [FREE Full text] [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
66. Willis EA, Szabo-Reed AN, Ptomey LT, Steger FL, Honas JJ, Al-Hihi EM, et al. Distance learning strategies for weight management utilizing online social networks versus group phone conference call. Obes Sci Pract 2017 Jun;3(2):134-142 [FREE Full text] [doi: 10.1002/osp4.96] [Medline: 28713582]
67. Lytle LA, Laska MN, Linde JA, Moe SG, Nanney MS, Hannan PJ, et al. Weight-gain reduction among 2-year college students: The CHOICES RCT. Am J Prev Med 2017 Feb;52(2):183-191 [FREE Full text] [doi: 10.1016/j.amepre.2016.10.012] [Medline: 27939237]
68. Gnagnarella P, Misotti AM, Santoro L, Akoumianakis D, del Campo L, de Lorenzo F, et al. Nutritional online information for cancer patients: a randomized trial of an internet communication plus social media intervention. J Cancer Educ 2016 Sep;31(3):472-480. [doi: 10.1007/s13187-015-0820-5] [Medline: 25820605]
69. Turner-McGrievy G, Tate D. Tweets, apps, and pods: results of the 6-month mobile pounds off digitally (mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011 Dec 20;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
70. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. Ann Intern Med 2013 Sep 3;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
71. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
72. Washington TA, Applewhite S, Glenn W. Using Facebook as a platform to direct young black men who have sex with men to a video-based HIV testing intervention: a feasibility study. Urban Soc Work 2017;1(1):36-52 [FREE Full text] [doi: 10.1891/2474-8684.1.1.36] [Medline: 29276800]
73. Cheung YT, Chan CH, Lai CJ, Chan WF, Wang MP, Li HC, et al. Using WhatsApp and Facebook online social groups for smoking relapse prevention for recent quitters: a pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015 Oct 22;17(10):e238 [FREE Full text] [doi: 10.2196/jmir.4829] [Medline: 26494159]
74. Stoddard JL, Augustson EM, Moser RP. Effect of adding a virtual community (bulletin board) to smokefree.gov: randomized controlled trial. J Med Internet Res 2008 Dec 19;10(5):e53 [FREE Full text] [doi: 10.2196/jmir.1124] [Medline: 19097974]
75. Chai W, Zou G, Shi J, Chen W, Gong X, Wei X, et al. Evaluation of the effectiveness of a WHO-5A model based comprehensive tobacco control program among migrant workers in Guangdong, China: a pilot study. BMC Public Health 2018 Feb 27;18(1):296 [FREE Full text] [doi: 10.1186/s12889-018-5182-6] [Medline: 29486753]
76. Graham AL, Cha S, Cobb NK, Fang Y, Niaura RS, Mushro A. Impact of seasonality on recruitment, retention, adherence, and outcomes in a web-based smoking cessation intervention: randomized controlled trial. J Med Internet Res 2013 Nov 7;15(11):e249 [FREE Full text] [doi: 10.2196/jmir.2880] [Medline: 24201304]
77. Graham AL, Cobb NK, Papandonatos GD, Moreno JL, Kang H, Tinkelman DG, et al. A randomized trial of Internet and telephone treatment for smoking cessation. Arch Intern Med 2011 Jan 10;171(1):46-53 [FREE Full text] [doi: 10.1001/archinternmed.2010.451] [Medline: 21220660]
78. Koufopoulos JT, Conner MT, Gardner PH, Kellar I. A web-based and mobile health social support intervention to promote adherence to inhaled asthma medications: randomized controlled trial. J Med Internet Res 2016 Jun 13;18(6):e122 [FREE Full text] [doi: 10.2196/jmir.4963] [Medline: 27298211]
79. Horvath KJ, Oakes JM, Rosser BR, Danilenko G, Vezina H, Amico KR, et al. Feasibility, acceptability and preliminary efficacy of an online peer-to-peer social support ART adherence intervention. AIDS Behav 2013 Jul;17(6):2031-2044 [FREE Full text] [doi: 10.1007/s10461-013-0469-1] [Medline: 23553347]
80. Greene J, Sacks R, Piniewski B, Kil D, Hahn JS. The impact of an online social network with wireless monitoring devices on physical activity and weight loss. J Prim Care Community Health 2013 Jul 1;4(3):189-194. [doi: 10.1177/2150131912469546] [Medline: 23799706]

81. Herring SJ, Cruice JF, Bennett GG, Darden N, Wallen JJ, Rose MZ, et al. Intervening during and after pregnancy to prevent weight retention among African American women. Prev Med Rep 2017 Sep;7:119-123 [FREE Full text] [doi: 10.1016/j.pmedr.2017.05.015] [Medline: 28660118]
82. Godino JG, Merchant G, Norman GJ, Donohue MC, Marshall SJ, Fowler JH, et al. Using social and mobile tools for weight loss in overweight and obese young adults (project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755 [FREE Full text] [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
83. West DS, Monroe CM, Turner-McGrievy G, Sundstrom B, Larsen C, Magradey K, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. J Med Internet Res 2016 Jun 13;18(6):e133 [FREE Full text] [doi: 10.2196/jmir.5474] [Medline: 27296086]
84. Daly N, Farren M, McKeating A, O Kelly R, Stapleton M, Turner MJ. A medically supervised pregnancy exercise intervention in obese women: a randomized controlled trial. Obstet Gynecol 2017 Nov;130(5):1001-1010. [doi: 10.1097/AOG.0000000000002267] [Medline: 29016485]
85. Nam Y. University of Southern California. 2014. The Effect of Twitter Social Support on Health Outcomes and Its Mediators: A Randomized Controlled Trial of a Social Support Intervention on Twitter For Patients Affected by Cancer URL: http://elearnica.ir/start/91cc91ba4de827d935f8c2d174275afa [accessed 2020-04-27]
86. Castillo RC, Wegener ST, Newell MZ, Carlini AR, Bradford AN, Heins SE, et al. Improving outcomes at level I trauma centers: an early evaluation of the trauma survivors network. J Trauma Acute Care Surg 2013 Jun;74(6):1534-1540. [doi: 10.1097/TA.0b013e3182921606] [Medline: 23694884]
87. Morris RR, Schueller SM, Picard RW. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. J Med Internet Res 2015 Mar 30;17(3):e72 [FREE Full text] [doi: 10.2196/jmir.4167] [Medline: 25835472]
88. Linden K, Sparud-Lundin C, Adolfsson A, Berg M. Well-being and diabetes management in early pregnant women with type 1 diabetes mellitus. Int J Environ Res Public Health 2016 Aug 22;13(8):e836 [FREE Full text] [doi: 10.3390/ijerph13080836] [Medline: 27556476]
89. Cobb NK, Poirier J. Effectiveness of a multimodal online well-being intervention: a randomized controlled trial. Am J Prev Med 2014 Jan;46(1):41-48 [FREE Full text] [doi: 10.1016/j.amepre.2013.08.018] [Medline: 24355670]
90. Namkoong K, Nah S, Record RA, van Stee SK. Communication, reasoning, and planned behaviors: unveiling the effect of interactive communication in an anti-smoking social media campaign. Health Commun 2017 Jan;32(1):41-50. [doi: 10.1080/10410236.2015.1099501] [Medline: 27119592]
91. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
92. Lau AY, Sintchenko V, Crimmins J, Magrabi F, Gallego B, Coiera E. Impact of a web-based personally controlled health management system on influenza vaccination and health services utilization rates: a randomized controlled trial. J Am Med Inform Assoc 2012;19(5):719-727 [FREE Full text] [doi: 10.1136/amiajnl-2011-000433] [Medline: 22582203]
93. Muntaner-Mas A, Vidal-Conti J, Borràs PA, Ortega FB, Palou P. Effects of a Whatsapp-delivered physical activity intervention to enhance health-related physical fitness components and cardiovascular disease risk factors in older adults. J Sports Med Phys Fitness 2017;57(1-2):90-102. [doi: 10.23736/S0022-4707.16.05918-1] [Medline: 26364690]
94. Owen J, Bantum E, Criswell K, Stanton A. Characterizing patterns of engagement with a socialnetworking intervention to treat cancer-related distress. Asia Pac J Clin Oncol 2012;8(Suppl 3):142-143 [FREE Full text]
95. Sun WH, Wong CK, Wong WC. A peer-led, social media-delivered, safer sex intervention for Chinese college students: randomized controlled trial. J Med Internet Res 2017 Aug 9;19(8):e284 [FREE Full text] [doi: 10.2196/jmir.7403] [Medline: 28793980]
96. Namkoong K, McLaughlin B, Yoo W, Hull SJ, Shah DV, Kim SC, et al. The effects of expression: how providing emotional support online improves cancer patients' coping strategies. J Natl Cancer Inst Monogr 2013 Dec;2013(47):169-174 [FREE Full text] [doi: 10.1093/jncimonographs/lgt033] [Medline: 24395987]
97. George DR, Dellasega C, Whitehead MM, Bordon A. Facebook-based stress management resources for first-year medical students: a multi-method evaluation. Comput Hum Behav 2013 May;29(3):559-562. [doi: 10.1016/j.chb.2012.12.008]
98. Moy ML, Janney AW, Nguyen HQ, Matthess KR, Cohen M, Garshick E, et al. Use of pedometer and internet-mediated walking program in patients with chronic obstructive pulmonary disease. J Rehabil Res Dev 2010;47(5):485-496 [FREE Full text] [doi: 10.1682/jrrd.2009.07.0091] [Medline: 20803392]
99. Bélanger-Gravel A, Godin G, Amireault S. A meta-analytic review of the effect of implementation intentions on physical activity. Health Psychol Rev 2013 Mar;7(1):23-54. [doi: 10.1080/17437199.2011.560095]
100. Fredrix M, McSharry J, Flannery C, Dinneen S, Byrne M. Goal-setting in diabetes self-management: a systematic review and meta-analysis examining content and effectiveness of goal-setting interventions. Psychol Health 2018 Aug;33(8):955-977. [doi: 10.1080/08870446.2018.1432760] [Medline: 29498547]
101. Kadushin C. Understanding Social Networks: Theories, Concepts, and Findings. Oxford, UK: Oxford University Press; 2012.

102. Southwell BG. Social Networks and Popular Understanding of Science and Health: Sharing Disparities. Baltimore, MD: Johns Hopkins University Press; 2013.
103. Lang T. Advancing global health research through digital technology and sharing data. Science 2011 Feb 11;331(6018):714-717. [doi: 10.1126/science.1199349] [Medline: 21311011]
104. Beran D, Byass P, Gbakima A, Kahn K, Sankoh O, Tollman S, et al. Research capacity building-obligations for global health partners. Lancet Glob Health 2017 Jun;5(6):e567-e568 [FREE Full text] [doi: 10.1016/S2214-109X(17)30180-8] [Medline: 28495256]
105. Ng LC, Hanlon C, Yimer G, Henderson DC, Fekadu A. Ethics in global health research: the need for balance. Lancet Glob Health 2015 Sep;3(9):e516-e517 [FREE Full text] [doi: 10.1016/S2214-109X(15)00095-9] [Medline: 26275322]
106. Eysenbach G, Consort-eHealth Group. Consort-eHealth: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

BCT: behavior change technique
BCTTv1: Behavior Change Technique Taxonomy version 1
MEDLINE: Medical Literature Analysis and Retrieval System Online, or MEDLARS Online
RCT: randomized controlled trials
SWAR: study within a review
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.08.19; peer-reviewed by S Thiebes, E Gabarron; comments to author 07.10.19; revised version received 26.11.19; accepted 22.03.20; published 11.06.20.

Please cite as:
Simeon R, Dewidar O, Trawin J, Duench S, Manson H, Pardo Pardo J, Petkovic J, Hatcher Roberts J, Tugwell P, Yoganathan M, Presseau J, Welch V
Behavior Change Techniques Included in Reports of Social Media Interventions for Promoting Health Behaviors in Adults: Content Analysis Within a Systematic Review
J Med Internet Res 2020;22(6):e16002
URL: http://www.jmir.org/2020/6/e16002/
doi:10.2196/16002
PMID:32525482

Original Paper

A Social Media Study on the Associations of Flavored Electronic Cigarettes With Health Symptoms: Observational Study

Long Chen[1]; Xinyi Lu[1]; Jianbo Yuan[1], PhD; Joyce Luo[2]; Jiebo Luo[1], PhD; Zidian Xie[3], PhD; Dongmei Li[3], PhD

[1]University of Rochester, Rochester, NY, United States
[2]Princeton University, Princeton, NJ, United States
[3]University of Rochester Medical Center, Rochester, NY, United States

Corresponding Author:
Dongmei Li, PhD
University of Rochester Medical Center
265 Crittenden Boulevard CU 420708
Rochester, NY, 14642-0708
United States
Phone: 1 5852767285
Email: Dongmei_Li@urmc.rochester.edu

Abstract

Background: In recent years, flavored electronic cigarettes (e-cigarettes) have become popular among teenagers and young adults. Discussions about e-cigarettes and e-cigarette use (vaping) experiences are prevalent online, making social media an ideal resource for understanding the health risks associated with e-cigarette flavors from the users' perspective.

Objective: This study aimed to investigate the potential associations between electronic cigarette liquid (e-liquid) flavors and the reporting of health symptoms using social media data.

Methods: A dataset consisting of 2.8 million e-cigarette–related posts was collected using keyword filtering from Reddit, a social media platform, from January 2013 to April 2019. Temporal analysis for nine major health symptom categories was used to understand the trend of public concerns related to e-cigarettes. Sentiment analysis was conducted to obtain the proportions of positive and negative sentiment scores for all reported health symptom categories. Topic modeling was applied to reveal the topics related to e-cigarettes and health symptoms. Furthermore, generalized estimating equation (GEE) models were used to quantitatively measure potential associations between e-liquid flavors and the reporting of health symptoms.

Results: Temporal analysis showed that the Respiratory category was consistently the most discussed health symptom category among all categories related to e-cigarettes on Reddit, followed by the Throat category. Sentiment analysis showed higher proportions of positive sentiment scores for all reported health symptom categories, except for the Cancer category. Topic modeling conducted on all health-related posts showed that 17 of the top 100 topics were flavor related. GEE models showed different associations between the reporting of health symptoms and e-liquid flavor categories, for example, lower association of the Beverage flavors with Respiratory compared with other flavors and higher association of the Fruit flavors with Cardiovascular than other flavors.

Conclusions: This study identified different potential associations between e-liquid flavors and the reporting of health symptoms using social media data. The results of this study provide valuable information for further investigation of the health effects associated with different e-liquid flavors.

(J Med Internet Res 2020;22(6):e17496) doi:10.2196/17496

KEYWORDS

e-cigarette; social media; eHealth

Introduction

Background

Electronic cigarette (e-cigarette) is an electronic nicotine delivery system (ENDS) that vaporizes electronic cigarette liquid (e-liquid), a solution that contains a number of substances [1]. E-cigarette is promoted as a tobacco substitute, which usually delivers nicotine, flavorings, and other additives via an inhaled aerosol. Since the first device publicly marketed in 2003, the e-cigarette market has experienced a significant growth in prevalence among all age groups [2].

The flavored e-cigarette market has grown extremely rapidly and has become increasingly popular among teenagers and young adults. A recent study identified 7764 distinct e-liquid flavors in 2014, and the number doubled to 15,586 in 2017 [3]. Adolescents and young adults prefer electronic cigarettes (e-cigarettes) over conventional tobacco because they are more affordable, accessible, convenient, and, especially, come with a variety of flavors [1]. A recent examination of the Population Assessment of Tobacco and Health (PATH) Wave 2 data found that more than 75% of youth e-cigarette users used a flavored e-cigarette with Fruit being the most popular e-cigarette flavor category, followed by the Candy/Sweet flavor category [4]. Nearly two-thirds of the adult e-cigarette users vape (ie, the behavior of inhaling vaporized aerosol) a flavored e-cigarette with Menthol or Mint as the most popular flavor category, followed by the Fruit flavor category [4]. The 2019 National Youth Tobacco Survey showed that current e-cigarette consumption among high school students increased from 1.5% in 2011 to 27.5% in 2019, with an estimated 72% of high school e-cigarette users using a flavored e-cigarette [5]. In addition, the number of e-cigarette users in middle school (grades 6-8) increased from 3.3% in 2017 to 10.5% in 2019, with nearly 60% of the middle school e-cigarette users using a flavored e-cigarette [5].

With the increasing popularity of e-cigarettes, it becomes important to understand the health symptoms associated with e-cigarettes and flavored e-cigarette consumption. There were claims that e-cigarettes are 95% safer than conventional tobacco, and nicotine released into the environment, if the e-cigarette liquid contains any, is negligible [6]. A previous study has shown that nonsmokers, such as children and pregnant women, had cardiovascular and other diseases due to passive vaping [7]. E-cigarette flavorings could damage human blood vessel cells in a laboratory environment, even in the absence of nicotine [8]. A number of studies have shown the association of vaping with some respiratory diseases or symptoms, including wheezing and asthma [9-12]. The cinnamon-flavored vaping products had the most potent cytotoxicity, leading to significantly decreased cell viability, increased reactive oxygen species levels, and other health symptoms [8]. Exposure to ENDS aerosol resulted in decreased metabolic activity and cell viability, whereas flavors significantly affected the toxicity of ENDS aerosol, with menthol, coffee, and strawberry flavors having the most significant impact on the overall cytotoxicity of e-cigarette products [13].

Objectives

E-cigarettes and e-cigarette use experiences are widely discussed on social media. Thus, the association of user-reported health risks with e-cigarette flavorings could be explored from social media posts. A recent review on e-liquid flavors using the JuiceDB data from 2013 to 2017 found that Fruity and Sweet were the two most popular e-liquid flavors and have more positive sentiments from users [14]. However, possible associations of e-liquid flavors with health risks have not been well investigated using social media data. Therefore, we aimed to examine probable associations between e-liquid flavors and the reporting of health symptoms using social media data through a data-driven approach.

Reddit was chosen as our main online data source. Reddit is a platform for any user (older than 13 years) to discuss, connect, and share their experiences and opinions online [15]. Besides not allowing the unwelcome content (eg, illegal, sexual, and confidential information), Reddit prohibit users from asking for votes or engaging in vote manipulation. As one of the biggest social media platforms in the United States, Reddit operates like a forum with divided communities called *subreddit*s, such that each one focuses on a specific topic such as sports and politics. Empirically, Reddit contains fewer advertisements and promotions but more self-reporting discussions from vapers (ie, people who vape), making it a desired source to reflect users' experiences and opinions [16]. In addition, discussions about e-liquid flavors have been shown to be one of the most prevalent topics about e-cigarettes on Reddit [17], with Fruit, Sweet, and Cream being the most prevalent flavors discussed among Reddit users [18]. A keyword list was generated by combining the keywords used in a previous e-cigarette–related social media study [19] and was used to filter out non-e-cigarette posts. To analyze the textual contents and users' opinions, topic modeling (Latent Dirichlet Analysis [LDA]) and sentiment analysis were conducted. We applied generalized estimating equation (GEE) models to quantify the potential associations between e-liquid flavors and major related health symptoms based on the e-cigarette–related discussions from Reddit users. Our study provides valuable information about the potential health effects of different e-liquid flavors, which could guide future research on flavored e-cigarettes.

Methods

Data Collection

We obtained the Reddit posts between January 01, 2013, and April 30, 2019, from pushshift.io [20]. We curated a list of e-cigarette–related terminologies (eg, vape, e-cig, e-liquid, and e-juice) from a prior study [19] on related topics as a preliminary list for keyword matching to generate an e-cigarette–related dataset. To account for typos and abbreviations, we applied Frequent Itemset Mining (FIM) based on the first iteration of the keyword-filtered dataset. FIM accepts all posts as input, and it outputs the combinations of most frequently appearing words in the dataset. We then manually selected high-frequency keywords to enrich the initial keyword list.

To investigate the e-cigarette–related health concerns discussed by Reddit users, we further generated a health-related subset (referred to as the health subset) from the e-cigarette dataset. A preliminary health keyword list was curated from a related prior study [21]. The list mostly comprised medical terminologies; thus, we applied FIM after filtering the preliminary list to include relevant conversational terms that frequently appear in our dataset. Similarly, a flavor-related subset (referred to as the flavor subset) was generated by using a flavor-related keyword list. The flavor keywords were generated by crawling the top e-cigarette manufacturers' websites to collect brand names, e-cigarette products' marketing names, and key ingredients.

All datasets were validated and denoised. We manually examined the noisy keywords (eg, blue, hard, and life can be flavor names or have other semantic meanings) from the keyword lists and removed the related posts if such keywords result in low precision during the filtering process. For keywords that were less semantically generic and noisy, we applied an additional filter to clean the data. For example, the flavor keyword "tobacco" contributes to a noisy flavor subset, which contains posts only about conventional tobacco products. We discovered a filter rule that filters out most of the undesired data while maintaining a decent level of recall to ensure the completeness of the dataset, by selecting 10 flavor-related keywords such as "flavor" and "e-liquid," which were found to frequently appear when users talk about tobacco-flavored e-cigarettes rather than conventional tobacco products. As a result, posts were classified as tobacco flavor-related when at least one sentence of the post contains both tobacco and one of the 10 flavor-related keywords. To examine the accuracy of our filter, we selected our top 10 flavor-related keywords, randomly picking 100 posts for each keyword. We then highlighted the sentences that contained the flavor-related keyword and manually examined whether the sentences were about discussions on e-cigarette–related topics. The accuracy of filtering was then calculated and is shown in Multimedia Appendix 1. Most flavor-related keywords have an accuracy between 0.8 and 0.9, which can be considered as clean.

Finally, we curated an e-cigarette keyword list of 20 terms ("e-cig," "e-cigs," "ecig," "ecigs," "electroniccigarette," "ecigarette," "ecigarettes," "vape," "vapers," "vaping," "vapes," "e-liquid," "ejuice," "eliquid," "e-juice," "vapercon," "vapeon," "vapefam," "vapenation," and "juul") and constructed the e-cigarette dataset consisting of 2,865,467 posts from 623,258 unique users. For data preprocessing, all URLs, email addresses, and non-English posts were removed to form a cleaner dataset. We created the health subset by applying 144 health keywords from nine health categories (Multimedia Appendix 1) and obtained 337,482 health-related posts from 138,448 unique users after filtering. Similarly, the flavor subset was generated with a flavor list of 1229 flavors (Multimedia Appendix 1) from 123 e-cigarette brands under 7 flavor categories including Fruit, Menthol or Mint, Tobacco, Sweet, Beverage, Mixed, and Other, containing 446,440 posts from 111,869 unique users.

Temporal Analysis

A temporal analysis was conducted to investigate longitudinal changes in the discussions about e-cigarettes on Reddit from January 2013 to April 2019. In addition, discussions about e-liquid flavors and health symptoms associated with e-cigarettes were examined using temporal analysis. A temporal trend for the percentage of Reddit posts for each reported health symptom was calculated as the number of Reddit posts mentioning each reported health symptom divided by the total number of e-cigarette related Reddit posts in each month.

Sentiment Analysis of Posts With Health Symptom Category Mentions

Sentiment analysis is a contextual analysis of sentences and paragraphs, which can extract subjective attitudes and opinions from the source documents. The Valence Aware Dictionary and sEntiment Reasoner (VADER) was used as the sentiment analyzer to extract users' opinions when discussing about health symptoms in each category on Reddit [22]. We calculated the sentiment score for each Reddit post and the average sentiment score for each health category. As some health-related keywords have negative sentiments (eg, headache), such keywords were replaced by a sentiment-neutral word "X." Such a process better reflects users' opinions about e-cigarettes. Sentiment propensity was then computed for each post with the suggested threshold from VADER. We classified posts with sentiment scores between +0.05 to +1.00 as positive posts, posts with sentiment scores between −0.05 to +0.05 (not including −0.05 and +0.05) as neutral, and posts with sentiment scores between −1.00 to −0.05 as negative. We normalized the numbers of positive, neutral, and negative posts by the total number of posts in each category to reflect the distribution of sentiment results. In addition, we compared the proportions of positive and negative posts within each health category using two proportion Z-tests in statistical analysis software R to determine the significant differences between the proportion of positive sentiments and the proportion of negative sentiments for each health category. We set the significance level for all tests at 5% and adjusted the original *P* values using the Bonferroni method to control for multiple testing error rate.

Topic Modeling

Topic modeling, specifically the LDA model, was applied to identify e-cigarette–related topics that were most frequently discussed [23]. LDA is a generative model for unsupervised topic modeling that can be regarded as a three-layer Bayesian distribution of words (or terms), topics, and documents such that each word in a document is expected to be allocated to a specific topic, where words in each topic are given a certain weight representing the possibility of appearance.

We applied topic modeling to the health subset. After data cleaning, all uppercase characters were converted to lowercase to ensure consistency, and words were lemmatized using spaCy to its stem form because different tenses will not be considered in the model. In addition, we removed stop-words (eg, you, I, and) using the Natural Language ToolKit.

As the LDA model considers each word's possibility of occurrence in a topic, we converted frequent phrases into one word so that a phrase can be considered as an *element* of a document, rather than a combination of separated words. For example, "throat" and "hit" are often mentioned together; thus, topics that contains both words as a "term" could be regarded as one element. We processed the dataset to find frequent bigrams (eg, panic attack, blood pressure, and lipid pneumonia) and trigrams (eg, food drug administration) using the Gensim package and then converted them into a single term.

Associations Between E-Liquid Flavors and Health Symptoms

To determine possible associations between e-liquid flavors and the reporting of health symptoms, we applied GEE models based on binomial distributions with logistic link functions. The within-user correlations were considered through the compound symmetry variance-covariance matrices under the GEE model framework. The GEE models were conducted using the statistical analysis software SAS v9.4 (SAS Institute Inc).

From the e-cigarette dataset, we selected posts with only one flavor-related keyword and at least one health-related keyword. If one selected post contained more than one health keyword, each health-flavor keyword pair was regarded as one unique entry. In total, 3336 entries were included in the dataset. From the GEE models, the probability of comentioning each flavor category with each health category was estimated, and the Tukey method [24] was used to adjust for multiplicity in pairwise comparisons between flavors.

The estimated probabilities of associations represent the probabilities that one flavor category and one health symptom category were comentioned by the same user. By performing pairwise comparisons, significant differences in the probabilities of association were identified within each health category. A heatmap was created to highlight the probabilities of associations between e-cigarette flavors and the reporting of health symptoms. All tests were two-sided with a significance level of 5%.

Results

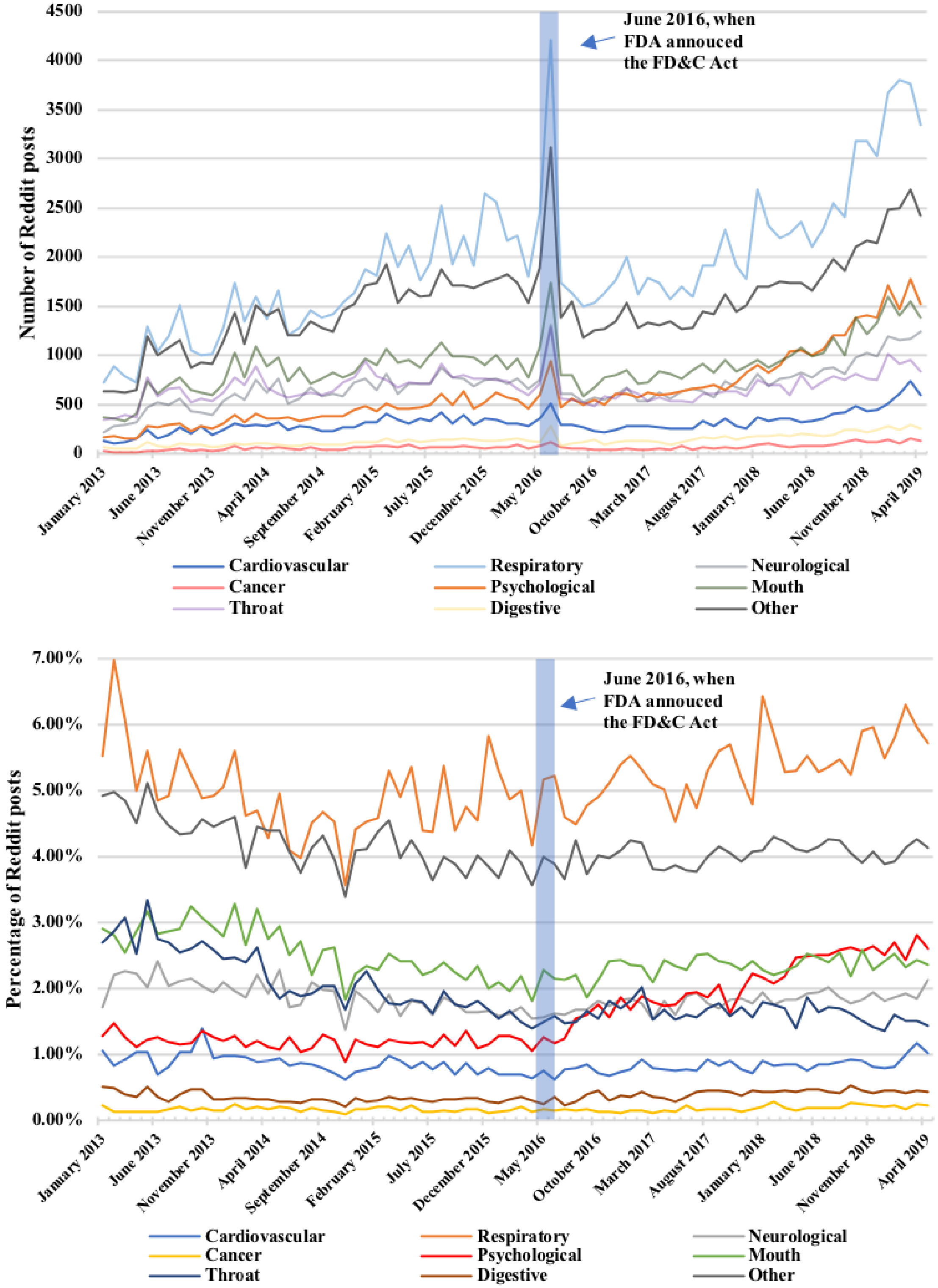
Classification of Health Symptoms

To better understand the potential association of e-liquid flavors with the reporting of health symptoms, we grouped 144 health keywords collected from Reddit into nine health categories, such as Respiratory, Neurological, and Cardiovascular (Multimedia Appendix 1). By counting the number of times mentioned in 337,482 Reddit posts, we found that Respiratory was the most mentioned health category (94,691/337,482, 28.06%), followed by Other (75,102/337,482, 22.25%), Mouth (44,212/337,482, 13.10%), Neurological (33,650/337,482, 9.97%), Throat (33,151/337,482, 9.82%), Psychological (31,041/337,482, 9.20%), Cardiovascular (15,460/337,482, 4.58%), and Digestive (6876/337,482, 2.04%) categories. Cancer was the least mentioned health category on our Reddit dataset at only 0.97% (3299/337,482).

Temporal Analysis of Health Symptom Category Mentioned on Reddit

To determine whether the frequency of each health symptom category mentioned on Reddit changes over time, we examined their longitudinal trend. Our results showed that the number of e-cigarette–related Reddit posts were steadily increasing over time since 2013 (Multimedia Appendix 2). We observed a spike in the temporal trend in June 2016 due to an increase in community discussion around the release of the Federal Food, Drug and Cosmetic (FD&C) Act on tobacco products and its substitutes, which prohibited e-cigarette and peripheral products to be sold to people aged under 18 years [7]. By examining the frequency of each health symptom category mentioned on Reddit over time, we observed that each health symptom category has a similar uptrend (Figure 1), which is similar to all e-cigarette–related posts. We observed that the Respiratory category mentions were consistently the highest among all health categories, whereas the Cancer category was the lowest. Spikes were also observed in the unnormalized temporal trend in June 2016 for all health categories due to the increased discussion on the FD&C Act. After normalization to the number of e-cigarette–related posts in each month, the percentage of health-related keywords mentions was relatively consistent over time, with one exception that the Psychological category had a noticeable growth from June 2016 to April 2019 (Figure 1).

Figure 1. Monthly health-related posts count and proportion of health-related posts (normalized to electronic cigarette–related posts) from January 2013 to April 2019. The blue, shaded region indicates the period where the US Food & Drug Administration (FDA) released the 2016 Federal Food, Drug and Cosmetic Act (FD&C Act).



Sentiment Analysis of Posts With Mentions of Health Categories

Sentiment analysis was conducted applied to the health-related subset posts to understand users' opinions regarding health symptoms in the discussion of e-cigarettes. Most health categories have a positive average sentiment score, whereas the Cancer category yields a negative average sentiment score (Table 1). For most health categories, the proportions of positive sentiment posts were significantly higher than the proportions of negative sentiment posts (*P*<.001), except for the Cancer category (Multimedia Appendix 3). Posts in the Cancer category had mostly negative sentiments, with over 70% of the posts identified with negative sentiment scores (*P*<.001). Example posters used in sentiment analysis are shown in Multimedia Appendix 1. For all categories, significantly more positive and negative sentiment posts were identified in comparison with neutral posts, suggesting that posts on Reddit usually had polarized opinions.

Table 1. Total post count and average sentiment score of posts in health categories.

Health category	Total post count	Average sentiment score
Respiratory	93,259	0.197
Cardiovascular	20,111	0.1889
Neurological	49,351	0.2862
Psychological	41,454	0.2104
Digestive	9123	0.1296
Mouth	62,011	0.2186
Throat	36,250	0.3062
Cancer	4983	−0.333

Potential Associations Between E-liquid Flavors and Health Categories

The LDA topic model was applied to the health-related subset posts to reveal content-wise insights from the discussions on e-cigarette–related health symptoms. As we used the health subset, the majority of the contents were health related. Noticeably, 17 out of 100 topics were mostly related to e-liquid flavors (Textbox 1). A number of representative posts are also provided for positive and negative sentiment posts in each health category to show the topics of discussions on Reddit using a qualitative approach (Multimedia Appendix 1).

Textbox 1. Examples of top 100 topics observed from health-related posts on Reddit.

Health related (n=7)

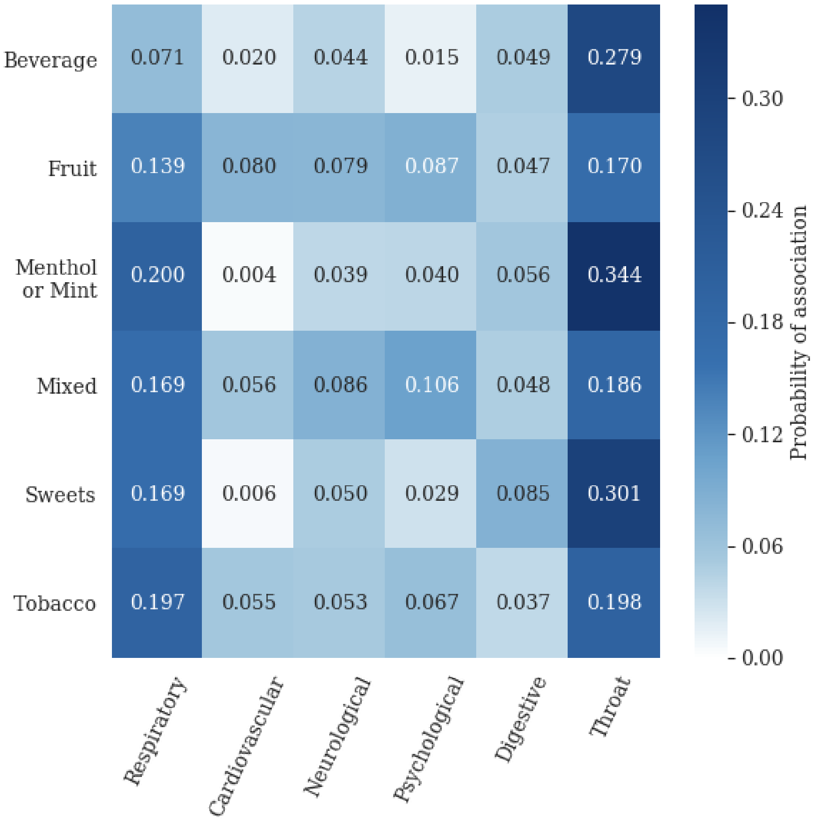
- Pain, symptom, relief, work, migraine, chronic, doctor, nausea, pill, and stomach
- Throat, vape, cold, allergy, irritate, sore, sensitive, sick, reaction, and irritation
- Smoking, vape, healthy, lung, alternative, bad, vaping, smoke, compare, and unhealthy
- Anxiety, depression, medication, attention-deficit/hyperactivity disorder, med, calm, panic attack, mental, symptom, and anxious
- Lung, cancer, die, kill, vape, death, tar, smoking, year, and chronic obstructive pulmonary disease
- Feel, vape, headache, sick, give, head, buzz, make, stomach, and feeling
- Cough, lung, asthma, breath, breathe, breathing, inhaler, clear, phlegm, and smoke

Flavor related (n=10)

- Juice, flavor, menthol, vendor, taste, order, vapor, bottle, sample, and Mt Baker Vapor, a major flavor concentrate wholesale for e-liquid DIY (MBV)
- Flavor, taste, strawberry, sweet, candy, favorite, fruit, juice, milk, and custard
- Vape, “flavour,” “vapour,” liquid, find, bit, UK, good, give, and taste
- Hit, throat, harsh, smooth, propylene glycol (PG), ratio, high, burn, vegetable glycerin, strong
- Flavor, taste, tongue, vaper, juice, strong, sweet, tasting, sense, and burn
- High, tetrahydrocannabinol, strain, cannabidiol, effect, anxiety, edible, cannabis, indica, and sativa
- Lung, popcorn, diacetyl, chemical, case, cigarette, inhale, amount, juice, and find
- Liquid, ingredient, propylene glycol, flavoring, safe, chemical, nicotine, e-liquid, PG, and flavor
- Flavor, taste, strawberry, sweet, candy, favorite, fruit, juice, milk, and custard
- Flavor, sweet, recipe, mix, taste, note, The Flavor Apprentice, a major flavor concentrate wholesale for e-liquid DIY (TFA), cream, cap, and candy

The probability of each health category being simultaneously mentioned with e-liquid flavors was calculated from GEE models to examine the potential associations of e-liquid flavors with the reporting of health symptoms. The associations between some e-liquid flavors and some health symptoms were not investigated due to small sample sizes, which left us with the associations between six e-liquid flavors and six health symptoms (Figure 2). As shown in Figure 2, the Respiratory category had high probabilities of association with most flavors, including Menthol or Mint (0.200), Tobacco (0.197), Sweet (0.169), and Mixed (0.169). The Fruit flavor category (0.080) had a relatively higher probability of association with the Cardiovascular category than the other flavor categories. Both the Fruit (0.079) and Mixed (0.086) flavor categories had a relatively higher probability of association with the Neurological category than other flavor categories. The Mixed flavor (0.106) category had a relatively higher probability of association with the Psychological category than other flavors. The Sweet flavor (0.085) category had a relatively higher probability of association with the Digestive category than other flavors. The Throat category had relatively high probabilities of association with all flavor categories, with Menthol or Mint (0.344), Sweet (0.301), and Beverage (0.279) being the highest.

Figure 2. Heatmap of the estimated probabilities of associations between e-liquid flavors and health symptoms on Reddit.



To test whether any e-liquid flavors have significantly higher probabilities of association with a certain health category than other flavors, we performed pairwise comparisons among e-liquid flavor categories on their associations within each health category (Multimedia Appendix 4). The Tobacco flavor category had a significantly higher probability of association with Respiratory than both Fruit and Beverage flavors. Both the Menthol or Mint and Mixed flavor categories had significantly higher probabilities of association with the Respiratory category than Beverage. Fruit flavor had higher probabilities of association with the Cardiovascular category than Menthol or Mint and Sweet. In the Neurological category, Mixed flavor had a significantly higher probability of association than the Tobacco flavor. For the Psychological category, Mixed flavor had a significantly higher probability of association than other flavor categories, except for Fruit flavor. Sweet flavor had higher probabilities of association than tobacco in the Digestive category. In the Throat category, a number of flavor categories had significantly higher probabilities of association than other categories (Menthol or Mint over Fruit, Mixed, and Tobacco; Sweet over Mixed and Tobacco; and Beverage over Fruit). No significant pairwise comparisons were observed between e-liquid flavors for the Cancer or Mouth category.

Discussion

Principal Findings

By examining the frequency of health keywords mentioned on Reddit, we showed that Respiratory was the most dominant health category, followed by Mouth, Throat, and Neurological categories. A previous social media study showed that the Mouth and Throat (combined), Respiratory, and Neurological health categories were the most frequently mentioned health symptoms on Reddit [25]. These frequency differences could be a result of different health-related keywords used, as our study included more keywords (n=23) in the Respiratory category than the mentioned previous study (n=4). E-cigarette consumption could lead to decreased exhaled nitric oxide, increased respiratory impedance, and increased flow respiratory resistance after short-term e-cigarette use, all of which are indications of immediate adverse physiological effects [26]. Adolescent e-cigarette users have increased rates of chronic bronchitis symptoms [9]. Using the PATH Wave 2 national survey data, e-cigarette use has been shown to be associated with wheezing and other related respiratory symptoms [10]. A recent cross-sectional study using the combined Behavioral Risk Factor

Surveillance System 2016 and 2017 national survey data showed a significant association between e-cigarette use and asthma in US adults who never smoked combustible cigarettes [11]. The dominant temporal trend of the Respiratory category shown in our results, along with the previously reported effects of e-cigarettes on respiratory symptoms, suggests that respiratory diseases are highly related to e-cigarette use.

Our study analyzed users' opinions on e-cigarettes from posts containing mentions of health keywords using sentiment analysis. Most health categories were dominated by positive sentiment posts, except that the Cancer category was dominated by negative sentiment posts. These results indicated a generally positive opinion toward e-cigarettes when Reddit users mentioned health-related keywords during discussion. As a forum consists of mostly teenagers and young adults, most users seem to advocate novel products such as e-cigarettes. This demographic characteristic could contribute to the average positive sentiments. In terms of the average negative sentiment on the Cancer category, we found that most posts discussed about the anecdotes of relatives or friends contracting cancer in the past with consumption of conventional tobacco products. This is reasonable considering the relatively young age of majority of the Reddit users, as cancer is not a prevalent disease in this age group (age: 13 to 32 years) [27]. We also noticed that a great proportion of posters were comparing e-cigarettes with conventional tobacco products, most of whom argued about the benefits of the former over the latter, thus contributing to the rather positive sentiment of most health keyword categories. A previous investigation on health-related effects reported by e-cigarette users in online forums also identified positive effects on the respiratory system, which is consistent with our study [21]. The sentiment analyses from this study are consistent with the sentiment analyses from a previous study using Reddit posts, which found a 60.7% provaping rating using textual analysis [17]. A recent investigation on online forum posts related to the health effects associated with e-cigarettes from 2008 to 2015, however, found that negative sentiment posts were dominant in almost all health categories [28]. This discrepancy might be due to the differences in the time frame, sentiment analysis methods, and data sources. Overall, the average positive sentiment scores across most categories reflect the provaping attitude of teenagers and young adults.

This study investigated the potential associations between e-liquid flavors and the reporting of health symptoms. Other than the Mouth and Cancer categories, some e-liquid flavor categories were identified to have significantly higher association with some health symptom categories compared with other flavor categories. For example, the Mixed flavors had a higher association with the Psychological category than other flavors except Fruit, suggesting that the Mixed e-liquid flavors might lead to a higher risk of psychological symptoms than other flavors. Using monocytes, one previous study showed that mixed flavors had greater cytotoxicity and higher levels of reactive oxygen species than individual flavors, which suggests that mixed flavors might be more harmful to e-cigarette users in psychological symptoms [29]. For the Throat category, Menthol or Mint had higher associations than other flavors, which is consistent with the previous finding that Menthol or Mint flavors lead to strong throat hit. For the Respiratory category, Menthol or Mint and Tobacco flavors had higher associations than other flavors, but this needs further exploration through both epidemiological and experimental studies. Previous studies have shown that some flavoring chemicals such as acetoin (in buttery flavor) and maltol (in candy flavor) can induce inflammatory responses and impair epithelial barrier function in human bronchial epithelial cells [30]. It has been shown that diacetyl, contained in buttery or creamy e-liquid [31,32], might be associated with respiratory disorders [33,34]. Cinnamaldehyde in cinnamon-flavored e-liquid has been shown to correlate with respiratory diseases and lung irritations [35-37]. In addition, benzaldehyde in cherry flavor and furfural in sweet flavor might lead to the irritation of respiratory airways [38]. Our results also showed a high probability of association (0.14-0.17) between Fruit, Mixed, and Sweet flavors and respiratory symptoms, which is consistent with previous studies. Although these associations between e-liquid flavors and health categories require further clinical validation, our results provided a systematic investigation on the association of e-liquid flavors and health symptoms using social media data and some valuable guidance for further studies.

Limitations

Due to the characteristics of social media platforms, the user demographic information (including age, gender, etc.) as well as other factors associated with e-cigarettes such as Propylene Glycol/Vegetable Glycerin ratio and nicotine concentrations were not included in our data analysis. In addition, the Reddit data provide information from users' perspectives, not from controlled laboratory experiments or clinical trials. Although such data provide us with the first-hand experience of e-cigarette users, they need to be further validated by well-designed laboratory or clinical studies. The measures including e-liquid flavors and health symptoms in this study were just mentions by Reddit users, which did not necessarily mean that the users were using these e-liquid flavors or having these health symptoms. In addition, as e-cigarettes have become increasingly popular in recent years, the research field in e-cigarettes has rapidly evolved. More health effects related to e-cigarettes have been identified. The new research findings might affect people's opinions on e-cigarettes and their association with health symptoms. Meanwhile, a scare story related to e-cigarettes might also cause a flurry of health posts. This analysis did not consider the effects of this evolvement, which we will explore in our future studies.

Conclusions

Using social media data from Reddit, our study quantitatively measured potentially different associations between e-liquid flavors and health symptoms. Temporal analysis revealed that mentions of keywords in most health categories were increasing, as the Psychological category had the most significant increase. Sentiment analysis showed that most health categories have a positive average sentiment score, whereas the Cancer category yields a negative average sentiment score. This study also investigated the potential associations between the reporting of health symptoms and e-liquid flavors. Other than the Mouth and Cancer categories, some e-liquid flavor categories were identified to have a significantly higher association with some health symptom categories compared with other flavor categories. With e-cigarette being prevalent worldwide and more e-liquid flavors being available, the health risks associated with e-cigarettes with different flavors should be further investigated. The results from this study could provide guidelines for future clinical and social media studies on the potential associations between e-liquid flavors and health symptoms as well as valuable guidance in future research on flavored e-cigarettes.

Acknowledgments

This research was supported by the National Cancer Institute of the National Institutes of Health and the Food and Drug Administration Center for Tobacco Products under Award Number U54CA228110.

Authors' Contributions

LC, ZX, and DL conceived and designed the study. LC and XL analyzed the data. LC wrote the manuscript. JY, JoL, JiL, ZX, and DL assisted with interpretation of analyses and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental Tables.
[DOCX File , 24 KB - jmir_v22i6e17496_app1.docx]

Multimedia Appendix 2
Monthly e-cigarette-related Reddit posts count from January 2013 to April 2019.
[PNG File , 295 KB - jmir_v22i6e17496_app2.png]

Multimedia Appendix 3
Proportion of positive, neutral and negative sentiment posts in each health categories.
[PNG File , 71 KB - jmir_v22i6e17496_app3.png]

Multimedia Appendix 4
Probabilities of the association between e-cigarette flavors and health categories. Error bars represent the upper halve of 95% CI.
[PNG File , 202 KB - jmir_v22i6e17496_app4.png]

References

1. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. e-Cigarette Use Among Youth and Young Adults: A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 2016.
2. Rachel G, Neal B, Stanton G. Center for Tobacco Control Research and Education. 2013. Background Paper on E-cigarettes (Electronic Nicotine Delivery Systems) URL: https://escholarship.org/uc/item/13p2b72n [accessed 2020-05-07]
3. Hsu G, Sun J, Zhu S. Evolution of electronic cigarette brands from 2013-2014 to 2016-2017: analysis of brand websites. J Med Internet Res 2018 Mar 12;20(3):e80 [FREE Full text] [doi: 10.2196/jmir.8550] [Medline: 29530840]
4. Schneller LM, Bansal-Travers M, Goniewicz ML, McIntosh S, Ossip D, O'Connor RJ. Use of flavored electronic cigarette refill liquids among adults and youth in the US-results from wave 2 of the population assessment of tobacco and health study (2014-2015). PLoS One 2018;13(8):e0202744 [FREE Full text] [doi: 10.1371/journal.pone.0202744] [Medline: 30138412]
5. Cullen KA, Gentzke AS, Sawdey MD, Chang JT, Anic GM, Wang TW, et al. E-cigarette use among youth in the United States, 2019. J Am Med Assoc 2019 Nov 5:- epub ahead of print. [doi: 10.1001/jama.2019.18387] [Medline: 31688912]
6. Bond CM. Do the benefits of electronic cigarettes outweigh the risks? Can J Hosp Pharm 2018;71(1):44-47 [FREE Full text] [Medline: 29531397]
7. Qasim H, Karim ZA, Rivera JO, Khasawneh FT, Alshbool FZ. Impact of electronic cigarettes on the cardiovascular system. J Am Heart Assoc 2017 Aug 30;6(9):- [FREE Full text] [doi: 10.1161/JAHA.117.006353] [Medline: 28855171]
8. Lee WH, Ong S, Zhou Y, Tian L, Bae HR, Baker N, et al. Modeling cardiovascular risks of e-cigarettes with human-induced pluripotent stem cell-derived endothelial cells. J Am Coll Cardiol 2019 Jun 4;73(21):2722-2737. [doi: 10.1016/j.jacc.2019.03.476] [Medline: 31146818]

9. McConnell R, Barrington-Trimis JL, Wang K, Urman R, Hong H, Unger J, et al. Electronic cigarette use and respiratory symptoms in adolescents. Am J Respir Crit Care Med 2017 Apr 15;195(8):1043-1049 [FREE Full text] [doi: 10.1164/rccm.201604-0804OC] [Medline: 27806211]
10. Li D, Sundar IK, McIntosh S, Ossip DJ, Goniewicz ML, O'Connor RJ, et al. Association of smoking and electronic cigarette use with wheezing and related respiratory symptoms in adults: cross-sectional results from the population assessment of tobacco and health (PATH) study, wave 2. Tob Control 2020 Mar;29(2):140-147. [doi: 10.1136/tobaccocontrol-2018-054694] [Medline: 30760629]
11. Osei AD, Mirbolouk M, Orimoloye OA, Dzaye O, Uddin SM, Dardari ZA, et al. The association between e-cigarette use and asthma among never combustible cigarette smokers: behavioral risk factor surveillance system (BRFSS) 2016 & 2017. BMC Pulm Med 2019 Oct 16;19(1):180 [FREE Full text] [doi: 10.1186/s12890-019-0950-3] [Medline: 31619218]
12. Perez MF, Atuegwu NC, Oncken C, Mead EL, Mortensen EM. Association between electronic cigarette use and asthma in never-smokers. Ann Am Thorac Soc 2019 Nov;16(11):1453-1456. [doi: 10.1513/AnnalsATS.201904-338RL] [Medline: 31404509]
13. Leigh NJ, Lawton RI, Hershberger PA, Goniewicz ML. Flavourings significantly affect inhalation toxicity of aerosol generated from electronic nicotine delivery systems (ENDS). Tob Control 2016 Nov;25(Suppl 2):ii81-ii87 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053205] [Medline: 27633767]
14. Li Q, Wang C, Liu R, Wang L, Zeng DD, Leischow SJ. Understanding users' vaping experiences from social media: initial study using sentiment opinion summarization techniques. J Med Internet Res 2018 Aug 15;20(8):e252 [FREE Full text] [doi: 10.2196/jmir.9373] [Medline: 30111530]
15. Reddit. Reddit Content Policy URL: https://www.redditinc.com/policies/content-policy [accessed 2019-11-04]
16. Reddit. URL: https://www.redditinc.com/ [accessed 2019-11-04]
17. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]
18. Wang L, Zhan Y, Li Q, Zeng D, Leischow S, Okamoto J. An examination of electronic cigarette content on social media: analysis of e-cigarette flavor content on Reddit. Int J Environ Res Public Health 2015 Nov 20;12(11):14916-14935 [FREE Full text] [doi: 10.3390/ijerph121114916] [Medline: 26610541]
19. Lienemann BA, Unger JB, Cruz TB, Chu KH. Methods for coding tobacco-related Twitter data: a systematic review. J Med Internet Res 2017 Mar 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]
20. pushshift.io. URL: https://www.pushshift.io [accessed 2020-05-07]
21. Hua M, Alfi M, Talbot P. Health-related effects reported by electronic cigarette users in online forums. J Med Internet Res 2013 Apr 8;15(4):e59 [FREE Full text] [doi: 10.2196/jmir.2324] [Medline: 23567935]
22. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. 2015 Presented at: AAAI'1; June 1-4, 2014; Ann Arbor, Michigan, USA URL: https://www.researchgate.net/publication/275828927_VADER_A_Parsimonious_Rule-based_Model_for_Sentiment_Analysis_of_Social_Media_Text
23. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3(4-5):993-1002. [doi: 10.5555/944919.944937]
24. Weed LL. Comparing individual means in the analysis of variance. Biometrics 1949 Jun;5(2):99-114. [Medline: 18151955]
25. Li Q, Zhan Y, Wang L, Leischow SJ, Zeng DD. Analysis of symptoms and their potential associations with e-liquids' components: a social media study. BMC Public Health 2016 Jul 30;16:674 [FREE Full text] [doi: 10.1186/s12889-016-3326-0] [Medline: 27475060]
26. Vardavas CI, Anagnostopoulos N, Kougias M, Evangelopoulou V, Connolly GN, Behrakis PK. Short-term pulmonary effects of using an electronic cigarette: impact on respiratory flow resistance, impedance, and exhaled nitric oxide. Chest 2012 Jun;141(6):1400-1406. [doi: 10.1378/chest.11-2443] [Medline: 22194587]
27. Mediakix: Influencer Marketing Agency. The Top 8 Reddit Statistics on Users, Demographics & More URL: https://mediakix.com/blog/reddit-statistics-users-demographics/ [accessed 2019-12-06]
28. Hua M, Sadah S, Hristidis V, Talbot P. Health effects associated with electronic cigarette use: automated mining of online forums. J Med Internet Res 2020 Jan 3;22(1):e15684 [FREE Full text] [doi: 10.2196/15684] [Medline: 31899452]
29. Muthumalage T, Prinz M, Ansah KO, Gerloff J, Sundar IK, Rahman I. Inflammatory and oxidative responses induced by exposure to commonly used e-cigarette flavoring chemicals and flavored e-liquids without nicotine. Front Physiol 2017;8:1130 [FREE Full text] [doi: 10.3389/fphys.2017.01130] [Medline: 29375399]
30. Gerloff J, Sundar IK, Freter R, Sekera ER, Friedman AE, Robinson R, et al. Inflammatory response and barrier dysfunction by different e-cigarette flavoring chemicals identified by gas chromatography-mass spectrometry in e-liquids and e-vapors on human lung epithelial cells and fibroblasts. Appl In Vitro Toxicol 2017 Mar 1;3(1):28-40 [FREE Full text] [doi: 10.1089/aivt.2016.0030] [Medline: 28337465]
31. Farsalinos K, Kistler K, Gillman G, Voudris V. Evaluation of electronic cigarette liquids and aerosol for the presence of selected inhalation toxins. Nicotine Tob Res 2015 Mar;17(2):168-174 [FREE Full text] [doi: 10.1093/ntr/ntu176] [Medline: 25180080]

32. Allen JG, Flanigan SS, LeBlanc M, Vallarino J, MacNaughton P, Stewart JH, et al. Flavoring chemicals in e-cigarettes: diacetyl, 2,3-pentanedione, and acetoin in a sample of 51 products, including fruit-, candy-, and cocktail-flavored e-cigarettes. Environ Health Perspect 2016 Jun;124(6):733-739 [FREE Full text] [doi: 10.1289/ehp.1510185] [Medline: 26642857]
33. McKernan L. Centers for Disease Control and Prevention. 2016. Criteria for a Recommended Standard: Occupational Exposure to Diacetyl and 2,3-Pentanedione URL: https://www.cdc.gov/niosh/docs/2016-111/default.html [accessed 2020-05-07]
34. Park H, O'Sullivan M, Vallarino J, Shumyatcher M, Himes BE, Park J, et al. Transcriptomic response of primary human airway epithelial cells to flavoring chemicals in electronic cigarettes. Sci Rep 2019 Feb 1;9(1):1400 [FREE Full text] [doi: 10.1038/s41598-018-37913-9] [Medline: 30710127]
35. Behar R, Davis B, Wang Y, Bahl V, Lin S, Talbot P. Identification of toxicants in cinnamon-flavored electronic cigarette refill fluids. Toxicol In Vitro 2014 Mar;28(2):198-208. [doi: 10.1016/j.tiv.2013.10.006] [Medline: 24516877]
36. Clapp PW, Pawlak EA, Lackey JT, Keating JE, Reeber SL, Glish GL, et al. Flavored e-cigarette liquids and cinnamaldehyde impair respiratory innate immune cell function. Am J Physiol Lung Cell Mol Physiol 2017 Aug 1;313(2):L278-L292 [FREE Full text] [doi: 10.1152/ajplung.00452.2016] [Medline: 28495856]
37. Clapp PW, Lavrich KS, van Heusden CA, Lazarowski ER, Carson JL, Jaspers I. Cinnamaldehyde in flavored e-cigarette liquids temporarily suppresses bronchial epithelial cell ciliary motility by dysregulation of mitochondrial function. Am J Physiol Lung Cell Mol Physiol 2019 Mar 1;316(3):L470-L486 [FREE Full text] [doi: 10.1152/ajplung.00304.2018] [Medline: 30604630]
38. Kaur G, Muthumalage T, Rahman I. Mechanisms of toxicity and biomarkers of flavoring and flavor enhancing chemicals in emerging tobacco and non-tobacco products. Toxicol Lett 2018 May 15;288:143-155 [FREE Full text] [doi: 10.1016/j.toxlet.2018.02.025] [Medline: 29481849]

Abbreviations

e-cigarettes: electronic cigarettes
e-liquid: electronic cigarette liquid
ENDS: electronic nicotine delivery system
FD&C: Federal Food, Drug and Cosmetic
FIM: Frequent Itemset Mining
GEE: generalized estimating equation
LDA: Latent Dirichlet Analysis
PATH: Population Assessment of Tobacco and Health
VADER: Valence Aware Dictionary and sEntiment Reasoner

Edited by G Eysenbach; submitted 16.12.19; peer-reviewed by P Talbot, R Shepherd; comments to author 26.01.20; revised version received 22.03.20; accepted 10.04.20; published 22.06.20.

Please cite as:
Chen L, Lu X, Yuan J, Luo J, Luo J, Xie Z, Li D
A Social Media Study on the Associations of Flavored Electronic Cigarettes With Health Symptoms: Observational Study
J Med Internet Res 2020;22(6):e17496
URL: http://www.jmir.org/2020/6/e17496/
doi:10.2196/17496
PMID:32568093

Original Paper

Association Between HIV-Related Tweets and HIV Incidence in the United States: Infodemiology Study

Robin Stevens[1], MPH, PhD; Stephen Bonett[1], BSN; Jacqueline Bannon[1], RN, PhD; Deepti Chittamuru[2], PhD; Barry Slaff[3], MSE, MCIT; Safa K Browne[4], MPH; Sarah Huang[1], BSN; José A Bauermeister[1], MPH, PhD

[1]Department of Family and Community Health, University of Pennsylvania School of Nursing, Philadelphia, PA, United States
[2]University of California Merced, Merced, CA, United States
[3]University of Pennsylvania, Philadelphia, PA, United States
[4]Children's Hospital of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Robin Stevens, MPH, PhD
Department of Family and Community Health
University of Pennsylvania School of Nursing
416 Curie Boulevard
Philadelphia, PA, 19104
United States
Phone: 1 2158984063
Email: robin.stevens@usc.edu

Abstract

Background: Adolescents and young adults in the age range of 13-24 years are at the highest risk of developing HIV infections. As social media platforms are extremely popular among youths, researchers can utilize these platforms to curb the HIV epidemic by investigating the associations between the discourses on HIV infections and the epidemiological data of HIV infections.

Objective: The goal of this study was to examine how Twitter activity among young men is related to the incidence of HIV infection in the population.

Methods: We used integrated human-computer techniques to characterize the HIV-related tweets by male adolescents and young male adults (age range: 13-24 years). We identified tweets related to HIV risk and prevention by using natural language processing (NLP). Our NLP algorithm identified 89.1% (2243/2517) relevant tweets, which were manually coded by expert coders. We coded 1577 HIV-prevention tweets and 17.5% (940/5372) of general sex-related tweets (including emojis, gifs, and images), and we achieved reliability with intraclass correlation at 0.80 or higher on key constructs. Bivariate and multivariate analyses were performed to identify the spatial patterns in posting HIV-related tweets as well as the relationships between the tweets and local HIV infection rates.

Results: We analyzed 2517 tweets that were identified as relevant to HIV risk and prevention tags; these tweets were geolocated in 109 counties throughout the United States. After adjusting for region, HIV prevalence, and social disadvantage index, our findings indicated that every 100-tweet increase in HIV-specific tweets per capita from noninstitutional accounts was associated with a multiplicative effect of 0.97 (95% CI [0.94-1.00]; P=.04) on the incidence of HIV infections in the following year in a given county.

Conclusions: Twitter may serve as a proxy of public behavior related to HIV infections, and the association between the number of HIV-related tweets and HIV infection rates further supports the use of social media for HIV disease prevention.

(J Med Internet Res 2020;22(6):e17196) doi:10.2196/17196

KEYWORDS

HIV/AIDS; social media; youth; natural language processing; surveillance

Introduction

The highest burden of new HIV infections has been reported in adolescents and young adults between the ages of 13 and 24 years, with 37.1% of the new HIV infections occurring in this age group in the United States [1]. Among the youths in this age group, 87% of the individuals diagnosed with HIV infection were reported to be young men, and 51% of these young men were identified as African American, while 25% of these young men were identified as Hispanic/Latino [1]. With the rapid increase in the usage of social media over the last 15 years, Twitter has emerged as a popular social networking platform. Studies have shown that Twitter is used by 32% of the adolescents and 44% of the young adults, with Black youths reporting higher levels of use than their white and Latino peers [2,3]. Since Twitter is used to discuss health-related and risk-related topics [4-6], this platform offers a distinct opportunity to investigate the attitudes, beliefs, and behaviors of the youths via their publicly shared posts that they have created or to which they have responded. This unique content may provide additional insights into the sentiments and discourses of youths [7] beyond what can be identified in traditional formative research methods, particularly at the national level. Analysis of Twitter, for example, might offer insight into the HIV-related beliefs and attitudes of youths of different races/ethnicities and help inform interventions that are designed to curb the HIV epidemic among youths.

The popularity of Twitter and the high volume of public tweets provide unprecedented access to discourses about sexual health and HIV by youths across a country. Although youths use social media platforms such as Twitter to share and seek sexual health information and to communicate with romantic and sexual partners [8], research on tweets related to alcohol, marijuana, cancer, and vaccines has shown that Twitter is also used to promote risky behaviors, spread misinformation, and reinforce HIV- and sexually transmitted infection (STI)-related stigmas [4,5,9,10]. Several studies have also considered social media messages as surveillance data to monitor the incidence of influenza, depression, Zika virus infections, and substance use [11-13]. Similar techniques have been used to assess the associations between social media messages on sex and HIV and the risk behavior and HIV incidence [14].

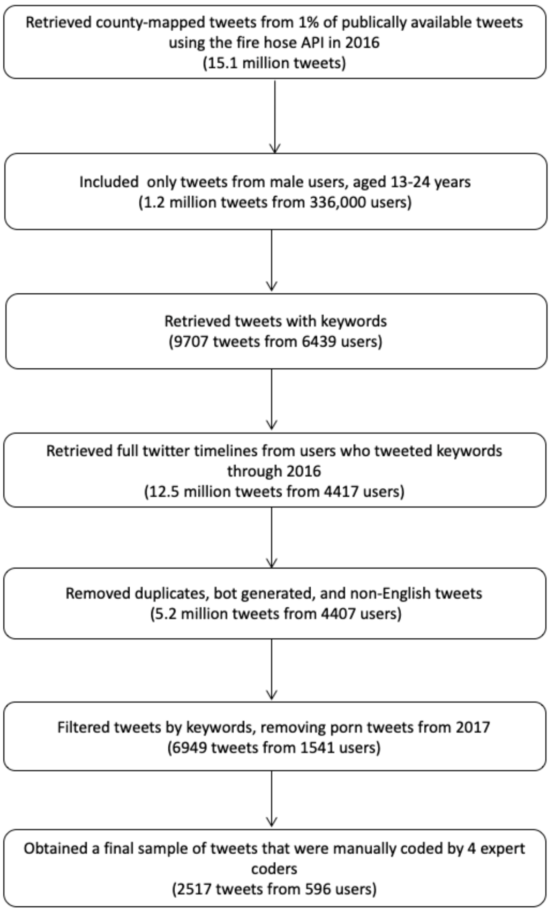
Several studies have shown evidence of a correlation between HIV-related tweets and HIV prevalence in a population [8,14,15]. Two studies [14,15] showed that future-oriented and action-based tweets regarding HIV were associated with decreased incidence of HIV infections at the county level in the United States. In contrast, Young et al [8] found that there was a statistically significant positive association between HIV-related tweets and HIV prevalence. However, these studies [8,14,15] did not distinguish between the source of the tweet; instead, they combined tweets from individual users and institutions such as public health agencies in their analyses. Importantly, studies on HIV and social media focus on certain keywords such as “sex,” “HIV testing,” and “discrimination.” Although the use of these keywords is useful for examining the associations between HIV-related tweets and HIV prevalence, studies often have reduced sensitivity to retrieve relevant tweets for analysis and intervention [16]. Since the abovementioned studies have provided promising evidence that tweets may be associated with HIV risk, there is a need for in-depth contextualized analysis of Twitter messages on risky sexual behavior and health, including analyses of message source variations. Therefore, the goal of our study was to explore how Twitter activity is related to HIV incidence and whether message characteristics such as content and source can reveal the incidence of HIV infection in a population and the future risks associated with HIV. Twitter messages may serve as a signal of the real-time dynamics in HIV epidemiology. In this study, we combined in-depth content analysis of HIV-related tweets with automated machine learning techniques to analyze the county-level associations between HIV-related tweets and new HIV infections in the United States.

Methods

Sample

Using the Twitter “fire hose” application programming interface, which provides broad access to public Twitter data, we drew a random sample of 1% of publicly available tweets posted between January 1, 2016 and December 31, 2016. We sampled tweets from users who tweeted at least 500 words in 2016 and who were geolocated in a county in the United States. To determine the geolocation, we used two types of data: tweet-specific latitude/longitude coordinates and the self-reported location information in Twitter user profiles. The distribution of the geolocated tweets by county approximate the US population density [17]. Duplicate tweets, bots, and non-English tweets were removed [18,19]. After we produced age and gender affiliation estimates for each user with HIV-related tweets, based on our tested algorithms [20,21], we limited our sample to users with predicted age (range, 13-24 years) and predicted gender (males only). Using previous literature [22] and input from our young researchers, we developed a keyword list of HIV-related terms such as HIV, AIDS, HIV testing, condoms, multiple sexual partners, STI, risky sexual behavior, and pre-exposure prophylaxis (PrEP). PrEP is an effective HIV prevention medication taken prior to exposure to the virus. This keyword list was used to identify relevant tweets, and we extracted 9707 HIV-related tweets from 6439 users from the age/gender stratified sample. We also removed pornographic tweets by developing a classifier to identify pornography and excluded those tweets from our data set. Our final data set included 6949 tweets by 1541 young male adults and male adolescents in the United States, and these tweets contained at least one relevant keyword. Figure 1 shows the number of messages and users retained at each step of the above described process.

Figure 1. Twitter sample retrieval flowchart. API: application programming interface.



Content Analysis

Data analysis was conducted with a sample of HIV-related tweets posted on Twitter. We manually coded a sample of HIV-related tweets by oversampling tweets with HIV-specific keywords. To accomplish this, we grouped our keywords into 2 broad categories: HIV prevention-specific tweets (n=1577) and general sex-related tweets (n=5372). We initially included a third category, namely, risk-related, which included risk behavior-promoting (n=6) tweets. However, we excluded this category from our analysis owing to the small number of tweets in this category. From the final data set, we took the full sample of 1577 prevention-related and a 17.5% (940/5372) random sample of general sex-related tweets, yielding 940 general sex-related tweets for manual content analysis.

The final data set (2517 tweets by 596 users) was coded by 4 expert coders for 19 nonexclusive categories. To capture the context, we expanded the coding unit beyond the initial tweet. The coders read the 5 tweets that proceeded from the tweet and the 5 tweets that followed the tweet. They also reviewed the images or webpages linked to the tweet. The coders achieved reliability by using a separate training data set of tweets, which was created through the same procedure used for the sample data set. We used a training set to train the coders without depleting the main data set. During training, the coders reconciled the differences in the code interpretations and coding approaches as a team. After the coding schema was finalized, the 4 coders achieved intercoder reliability on key constructs assessed with an intraclass correlation at 0.80 or higher. Approximately 20% (500/2517) of the final data set was coded by at least two coders.

Measures

The HIV incidence—the outcome variable—was assessed as the number of new cases of HIV infections in a given county in 2017. These data were sourced from the Centers for Disease Control and Prevention AtlasPlus data platform [23]. Counties with suppressed data owing to low case counts were assigned a value of 2, which represented the midpoint between the lowest possible suppressed value of 1 and the highest possible suppressed value of 4.

Twitter messages were classified into the following 3 categories in 2016: risk-specific Twitter activity, prevention-specific Twitter activity, and HIV-specific Twitter activity. Risk-specific Twitter activity is the sum of all the tweets categorized with a risk-related code (eg, multiple partners, pro risk-taking, substance use, transactional sex, and unprotected sex) in a given county, per 100,000 residents. Prevention-specific Twitter activity is the sum of all the tweets categorized as prevention-related (eg, antirisk taking, condoms, HIV testing, HIV/AIDS, PrEP, research, education, and news) in a given county, per 100,000 residents. HIV-specific Twitter activity is the sum of the risk-specific and prevention-specific Twitter activities, in addition to the tweets that were tagged as related to LGBTQ content. All users in our data set were identified as either an individual or an institution based on the manual review of the user profile and recent posting activity. Institutions included public health agencies, social service organizations, and advocacy groups, and typically included the organization name in the username or user description. Our final measures of the tweets consisted of 3 Twitter activity categories (risk, prevention, or HIV-specific) from individuals or institutions, resulting in 6 Twitter variables.

We accounted for 3 geographic control variables: HIV prevalence, social disadvantage, and census region. HIV prevalence in a geographic area is the key epidemiological factor linked to the number of new cases in that area [24]. We used county-level HIV prevalence rates in 2015 to account for the existing patterns of HIV infection. Counties with suppressed data owing to low case counts were assigned case counts of 6, which represented the midpoint between the lowest possible suppressed value of 1 and the highest possible suppressed value of 11. Studies have also shown that socioeconomic factors measured at the city-wide level are the key drivers of new HIV infections [25]. To capture the combined effect of multiple dimensions of socioeconomic disadvantage, we calculated the social disadvantage index at the county level for the counties in our study (Cronbach α=.82) [26]. This index was calculated by summing the z-scores for the percentage of the population living in poverty, the percentage of the population with a high school degree or equivalent, the median household income, and the percentage of the population lacking health insurance. These measures were obtained from the US Census Bureau Small Area Income and Poverty Estimates and the American Community Survey [27]. Negative weights were applied to high school education and median income, yielding an index that reflected greater social disadvantage for high values of the index and lesser social disadvantage for low values of the index. Census region was included to account for the regional variations in the HIV epidemic. The four regions, that is, northeast, south, Midwest, and west regions of the United States, were treated as the control variables in the models.

Statistical Analysis

General sex-related tweets were given sample weights of 6.25 for all the analyses to reflect the random samplings performed to reduce the data for coding. We used the Wilcoxon rank sum test with continuity correction for large samples to compare county tweet outputs based on the message source (ie, tweets from individuals vs institutions). We used negative binomial regression to estimate the effects of Twitter activity on HIV incidence at the county level. Our outcome of interest for this analysis was the rate of new diagnoses of HIV infections per capita at the county level. To model this rate variable, we included an offset term for the county population in 2017 in our regression analysis [28]. Negative binomial regression was chosen because our county outcome variables showed significant overdispersion from the Poisson distribution. Unadjusted models were run first for each of the 6 Twitter variables and the 3 control variables. Separate multivariate models were run for each of the Twitter variables, thereby adjusting for all the control variables. Variance inflation factors were examined for all final models, and none showed evidence of multicollinearity. Analyses were performed in R-3.5.1 [29] using the MASS package [30] glm.nb() function for negative binomial regression.

Results

Descriptive and Geospatial Data

Our data included 2517 tweets that were identified as potentially relevant to HIV risk (eg, unprotected sex) and prevention tags (eg, condom use, HIV testing, research, education), and these tweets originated in 109 counties across the United States. Of these, 940 were general sex-related tweets (including emojis, gifs, and images) and were given a sample weight to reflect our random sampling procedure. Each tweet in our data set represents 100 tweets in the real world as our data was drawn from 1% of publicly available tweets. However, we have reported all our results in units of true tweets, which were calculated by multiplying our results by 100. In 2016, 321 HIV-specific tweets, on an average, originated from individuals in each county. Counties had an average of 143 prevention-related and 118 risk-related tweets from individuals. An average of 944 HIV-specific tweets, 843 prevention-related tweets, and 31 risk-related tweets originating from institutions were sourced to each county. Institutions tweeted significantly more HIV-related (U=67,812; P<.001) and prevention-related messages (U=62,711; P<.001) and significantly less risk-related messages as compared to individuals (U=63,879; P<.001). Within counties that had at least one potentially relevant tweet, the median number of new HIV cases diagnosed in 2017 was 70 per county (range: 0-1530). HIV prevalence in these counties ranged from 6.02 to 2590 per 100,000 residents, with a median prevalence rate of 306 per 100,000 residents. The social disadvantage index ranged from –6.52 to 7.53 (Table 1).

The crude incidence rate ratios (IRRs) for each variable of interest of HIV incidence in 2017 were calculated using negative binomial regression with an offset for the county population (Table 2).

HIV prevalence in 2017 was positively associated with HIV prevalence in 2015 and social disadvantage index in 2015 (IRR 1.104, 95% CI 1.075-1.134; P<.001). Compared to that in the Midwest region, significantly higher HIV incidence was observed in the northeast (IRR 1.286, 95% CI 0.985-1.683; P<.001) and south (IRR 2.126, 95% CI 1.711-2.630; P<.001) regions of the United States. We did not observe a significant difference (IRR 0.967, 95% CI 0.749-1.250; P>.99) in the counties in the west region of the United States. The large number of prevention tweets from individuals in 2016 was significantly associated with the high incidence of HIV in the following year (IRR 1.082, 95% CI 1.003-1.183; P=.048). No other significant bivariate associations were found between HIV-related tweets and HIV incidence for combinations of tweet category and user type.

Table 1. Descriptive statistics at the county level (n=109).

Descriptive statistics	Values	
	Mean (SD)	Median (Min, Max)
HIV prevalence case count, 2017	173 (260)	70 (0.00, 1530)
HIV prevalence case count, 2015	484 (500)	306 (6.02, 2590)
County population, 2017	832,000 (1,090,000)	535,000 (13,900, 8,580,000)
Social disadvantage index, 2015	0.251 (3.06)	0.463 (–6.52, 7.53)

Table 2. Crude incidence rate ratios (bivariate models).[a]

Parameters	Crude incidence rate ratio	95% CI		*P* value
		Upper	Lower	
HIV tweets 2016, person	1.006	0.975	1.043	.65
Prevention 2016, person	1.082	1.003	1.183	.048
Risk tweets 2016, person	0.976	0.931	1.024	.23
HIV tweets 2016, institution	1.006	0.998	1.016	.13
Prevention tweets 2016, institution	1.006	0.997	1.018	.16
Risk tweets 2016, institution	1.155	0.876	1.651	.30
HIV prevalence, 2015	1.002	1.001	1.002	<.001
Social disadvantage index	1.104	1.075	1.134	<.001
Region of the United States				
Midwest	Ref[b]	Ref	Ref	Ref
Northeast	1.286	0.985	1.683	<.001
South	2.126	1.711	2.630	<.001
West	0.967	0.749	1.250	>.99

[a] All tweet variables are reported in units of 100 tweets.

[b] Ref: reference.

Multivariate Analyses

Multivariate models were used to test the adjusted effects for each of the 3 categories of tweets, for individuals and institutions separately, on HIV incidence in the following year. These models are summarized in Tables 3 and 4.

In all 6 models, HIV prevalence in 2017 was positively associated with HIV prevalence in 2015 and social disadvantage index. Additionally, all 6 models showed a significant difference in HIV incidence between the south and Midwest regions. Only one model, Model 1, showed statistically significant effect for a tweet variable on HIV incidence. In Model 1, HIV-specific tweets originating from individuals were negatively associated with HIV incidence at the county level in the following year, after adjusting for region, HIV prevalence, and social disadvantage index. Each additional 100 HIV-specific tweets per capita that originated from an individual in a given county was associated with a 3% decrease in the incidence rate of HIV in the following year, after adjusting for covariates.

Table 3. Multivariate models for tweets from individuals.[a]

	HIV incidence per capita, 2017					
	Model 1: HIV-specific tweets		Model 2: Prevention-specific tweets		Model 3: Risk-specific tweets	
Predictors	Incidence rate ratio (95% CI)	*P* value	Incidence rate ratio (95% CI)	*P* value	Incidence rate ratio (95% CI)	*P* value
HIV-specific Twitter activity, 2016	0.97 (0.94-1.00)	.04	N/A[b]	N/A	N/A	N/A
Prevention-specific Twitter activity, 2016	N/A	N/A	0.95 (0.90-1.01)	.13	N/A	N/A
Risk-specific Twitter activity, 2016	N/A	N/A	N/A	N/A	1.03 (0.86-1.24)	.73
HIV prevalence, 2015	1.00 (1.00-1.00)	<.001	1.00 (1.00-1.00)	<.001	1.00 (1.00-1.00)	<.001
Social disadvantage index	1.04 (1.02-1.06)	<.001	1.04 (1.02-1.06)	<.001	1.04 (1.02-1.06)	<.001
Census region						
Midwest	Ref[c]	Ref	Ref	Ref	Ref	Ref
North	0.90 (0.73-1.10)	.30	0.90 (0.74-1.10)	.29	0.90 (0.73-1.11)	.33
South	1.40 (1.18-1.67)	<.001	1.41 (1.20-1.67)	<.001	1.41 (1.18-1.68)	<.001
West	0.94 (0.78-1.15)	.55	0.95 (0.79-1.14)	.56	0.94 (0.77-1.14)	.51

[a] All tweet variables are reported in units of 100 tweets.

[b] N/A: not applicable.

[c] Ref: reference.

Table 4. Multivariate models for tweets from institutions.[a]

	HIV incidence per capita, 2017					
	Model 4: HIV-specific tweets		Model 5: Prevention-specific tweets		Model 6: Risk-specific tweets	
Predictors	Incidence rate ratios (95% CI)	*P* value	Incidence rate ratios (95% CI)	*P* value	Incidence rate ratios (95% CI)	*P* value
HIV-specific Twitter activity, 2016	1.00 (0.99-1.00)	.92	N/A[b]	N/A	N/A	N/A
Prevention-specific Twitter activity, 2016	N/A	N/A	1.00 (0.99-1.01)	.996	N/A	N/A
Risk-specific Twitter activity, 2016	N/A	N/A	N/A	N/A	1.03 (0.86-1.24)	.73
HIV prevalence, 2015	1.00 (1.00-1.00)	<.001	1.00 (1.00-1.00)	<.001	1.00 (1.00-1.00)	<.001
Social disadvantage index	1.04 (1.02-1.06)	<.001	1.04 (1.02-1.06)	<.001	1.04 (1.02-1.06)	<.001
Census region						
Midwest	Ref[c]	Ref	Ref	Ref	Ref	Ref
North	0.90 (0.73-1.11)	.34	0.90 (0.73-1.11)	.34	0.90 (0.73-1.11)	.33
South	1.41 (1.18-1.68)	<.001	1.41 (1.18-1.68)	<.001	1.41 (1.18-1.68)	<.001
West	0.94 (0.77-1.14)	.51	0.94 (0.77-1.14)	.51	0.94 (0.77-1.14)	.51

[a] All tweet variables are reported in units of 100 tweets.

[b] N/A: Not applicable.

[c] Ref: reference.

Discussion

Principal Findings

In this study, we analyzed the association between geolocated HIV-related tweets within the United States and the future incidence of HIV infection. HIV-specific tweets were more likely to emerge in those locations in the United States that had a high incidence of HIV. The number of HIV-specific tweets made by institution-associated accounts was higher than that of individual tweets. Interestingly, risk-related information in institution-associated tweets was lesser than that in tweets made by individual users. However, we did not observe significant associations between the number of HIV-specific tweets made by institutions and county-level HIV incidence. In contrast, increased numbers of HIV-specific tweets made by individual users were significantly associated with the decreased number of HIV cases in the following year at the county level, even when controlling for the geographic location. These findings suggest that the source of the tweet plays an important role, with individuals tweeting less about prevention, and these individual tweets showed a strong association with the future outcomes of HIV infections.

Geolocated conversations regarding HIV infections were negatively associated with county-level HIV incidence. These findings suggest that locations with few HIV-related Twitter posts and conversations by individuals may indicate those that require targeted interventions. Thus, counties with high incidence of HIV infections and few tweets may indicate an opportunity for increased investigation and potential intervention.

There are several possible reasons for our observation of low incidence of HIV infections in counties with large numbers of HIV-related tweets in the previous year. Increased numbers of HIV-related tweets at the county level could indicate increased community involvement, policy initiatives, and resource utilization in a given county [31]. Additionally, increased numbers of HIV-related tweets by individuals could reflect increased activities in addressing various determinants of HIV risks, including limited institutional support and reduced access to health care [32]. Although studies have sought to incorporate real-time analysis of Twitter data in association with localized HIV incidence, up-to-date HIV epidemiological data is limited or inaccessible to researchers. It is critical that public health professionals and computer scientists collaborate to develop novel approaches in analyzing Twitter data in accordance with the available HIV epidemiological data.

Our findings corroborate those of Ireland et al [15] but they are in contrast to those reported by Young et al [8] who found a positive association between HIV-related tweets and HIV prevalence. Although our study was similar in concept to that conducted by Young et al [8], we used a more specific definition of HIV-related tweets by excluding keywords that were less sensitive in our training sample (eg, “fuck”) and including a variety of slang terms that were compiled by our young researchers. Moreover, our study may have some slight differences from that of Young et al [8] because of the time period in our study—we may have identified a more recent phenomena in the prevalence of HIV infections. Two previous studies [14,15] analyzed the association between the HIV-related tweets and the corresponding epidemiological data in the same time period as considered in our study, whereas Young et al [8] analyzed the Twitter data with the epidemiological data of the previous year. The analyses in this study mirror those reported by Young et al [8] because we also analyzed the epidemiological data from the year after obtaining our specified frame of tweets in 2016.

Our findings suggest that discourses on HIV and risky sexual behavior on Twitter may serve as a signal of sexual health outcomes at the aggregate level. However, the low effect size and nonsignificant results of some of the models make it difficult to state this fact conclusively. It is clear that HIV-related discourse is geographically concentrated, and in coordination with epidemiological surveillance efforts, it may be used to inform intervention efforts. Despite the relative rarity of direct discussion of HIV on Twitter, this social media platform is still an important medium for conversations regarding HIV and health behaviors. Given its wide user base, Twitter can serve as a platform for discussing useful HIV prevention strategies, and such platforms deserve further investment as tools to end the HIV epidemic.

This study has several strengths. Our analysis combined NLP and manual coding, thereby allowing for coding of a large number of tweets for in-depth meaning, while preserving context. Our use of geolocated tweets allowed for location-based analysis with epidemiological and census data. These aspects of our study allowed for contextualized analysis of Twitter data, which may be useful for targeted interventions across the United States.

Limitations

This study has several limitations. First, we did not evaluate model significance by using multiple correction comparisons. Our study only used a set of geolocated tweets from 2016, which greatly reduced the available sample. This analysis included 10.8% (339/3141) of all the counties and there may have been different Twitter discourses in other regions that were not included in this analysis. It is possible that tweets that were not geolocated in that year could have revealed additional information about the nature of HIV-specific tweets relative to HIV incidence. Second, we excluded Spanish tweets, which limited our ability to capture the web-based discourse among Latino men. Third, since we focused on deidentified Twitter data, our study does not contain information on individual characteristics or behaviors.

Conclusion

With the increase of public discourse through Twitter, public health efforts leveraging this social medium are needed. Social media platforms such as Twitter offer an opportunity for health professionals to monitor population health and promote HIV disease prevention. We observed a negative association between HIV-specific tweets made by individual users and HIV incidence in the following calendar year at the county level. Our study underscores the importance of social media as a crucial aspect in the lives of individuals, as these discourses might unearth the youths' knowledge, attitudes, and beliefs related to HIV. Public health efforts seeking to use social media as a tool for HIV surveillance and intervention are warranted.

Acknowledgments

This work was supported by the National Institute of Health CFAR administrative supplement and the University of Pennsylvania Center of AIDS HIV Disparities Scientific Working Group. The authors would like to acknowledge Genevieve Afrifa, Obumneke Obi, Chinanuekpere Emma Ezeala, and Rashaan Galloway for their brilliant coding assistance.

Conflicts of Interest

None declared.

References

1. Centers FDC. https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2017-vol-29.pdf. Atlanta, GA: Centers for Disease Control and Prevention; 2017. HIV Surveillance Report, 2017 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2017-vol-29.pdf [accessed 2018-05-08]
2. Smith A. https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/. Washington, DC: Pew Research Center; 2018. Social media use in 2018 URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2019-01-31]
3. Perrin A, Anderson M. Facttank: News by the Numbers. Washington, DC: Pew Research Center; 2019 Apr 9. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018 URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-06-05]
4. Cabrera-Nguyen EP, Cavazos-Rehg P, Krauss M, Bierut LJ, Moreno MA. Young Adults' Exposure to Alcohol- and Marijuana-Related Content on Twitter. J Stud Alcohol Drugs 2016 Mar;77(2):349-353 [FREE Full text] [doi: 10.15288/jsad.2016.77.349] [Medline: 26997194]
5. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
6. Gabarron E, Serrano JA, Wynn R, Lau AY. Tweet content related to sexually transmitted diseases: no joking matter. J Med Internet Res 2014 Oct 06;16(10):e228 [FREE Full text] [doi: 10.2196/jmir.3259] [Medline: 25289463]
7. Kreniske P. Developing a culture of commenting in a first-year seminar. Computers in Human Behavior 2017 Jul;72:724-732. [doi: 10.1016/j.chb.2016.09.060]
8. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
9. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/ajph.2018.304567]
10. Chou WS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
11. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS One 2015;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
12. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/ajph.2016.303512]
13. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav 2017 Feb;65:289-295. [doi: 10.1016/j.addbeh.2016.08.019] [Medline: 27568339]
14. Ireland ME, Chen Q, Schwartz HA, Ungar LH, Albarracin D. Action Tweets Linked to Reduced County-Level HIV Prevalence in the United States: Online Messages and Structural Determinants. AIDS Behav 2016 Jun;20(6):1256-1264 [FREE Full text] [doi: 10.1007/s10461-015-1252-2] [Medline: 26650382]
15. Ireland ME, Schwartz HA, Chen Q, Ungar LH, Albarracín D. Future-oriented tweets predict lower county-level HIV prevalence in the United States. Health Psychol 2015 Dec;34S:1252-1260 [FREE Full text] [doi: 10.1037/hea0000279] [Medline: 26651466]
16. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
17. Schwartz H, Eichstaedt J, Kern M, Dziurzynski L, Lucas RE, Agrawal M, et al. Characterizing Geographic Variation in Well-Being Using Tweets. In: Proceedings of the 7th International AAAI Conference on Web and Social Media. 2013

Presented at: 7th International AAAI Conference on Web and Social Media; 7/8/2013; Cambridge, MA URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6138

18. Schwartz H, Giorgi S, Sap M, Crutchley P, Ungar L, Eichstaedt J. DLATK: Differential language analysis ToolKit. In: Proceedings of the 2017 EMNLP System Demonstrations. 2017 Sep 7 Presented at: EMNLP Systems Demonstrations; 2017; Copenhagen, Denmark p. 55-60 URL: https://wwbp.org/papers/DLATK_Differential_Language_Analysis_ToolKit.pdf [doi: 10.18653/v1/d17-2010]
19. Lui M, Baldwin T. langid.py: an off-the-shelf language identification tool. In: Proceedings of the ACL 2012 System Demonstrations.: Association for Computational Linguistics; 2012 Jul 1 Presented at: Association for Computer Linguistics '12; 2012-7-1; Stroudsburg, PA p. 25-30 URL: https://dl.acm.org/doi/10.5555/2390470.2390475
20. Sap M, Park G, Eichstaedt J, Kern M, Stillwell D, Kosinski M, et al. Developing Age and Gender Predictive Lexica over Social Media. In: http://emnlp2014.org/papers/pdf/EMNLP2014121.pdf.: Association for Computational Linguistics; 2014 Presented at: Conference on Empirical Methods in Natural Language Processing; 2014; Doha, Qatar p. 1146-1151 URL: http://emnlp2014.org/papers/pdf/EMNLP2014121.pdf
21. Preoiuc-Pietro D, Ungar L. User-Level Race and Ethnicity Predictors from Twitter Text. 2018 Presented at: Proceedings of the 27th International Conference on Computational Linguistics; 2018; Santa Fe, NM URL: https://www.aclweb.org/anthology/C18-1130/
22. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
23. Centers FDC. https://www.cdc.gov/nchhstp/atlas/index.htm. Atlanta, GA: Center for Disease Control and Prevention; 2017. NCHHSTP AtlasPlus URL: https://www.cdc.gov/nchhstp/atlas/index.htm [accessed 2018-02-02]
24. Hallett TB, Zaba B, Todd J, Lopman B, Mwita W, Biraro S, ALPHA Network. Estimating incidence from prevalence in generalised HIV epidemics: methods and validation. PLoS Med 2008 Apr 08;5(4):e80 [FREE Full text] [doi: 10.1371/journal.pmed.0050080] [Medline: 18590346]
25. Buot MG, Docena JP, Ratemo BK, Bittner MJ, Burlew JT, Nuritdinov AR, et al. Beyond race and place: distal sociological determinants of HIV disparities. PLoS One 2014;9(4):e91711 [FREE Full text] [doi: 10.1371/journal.pone.0091711] [Medline: 24743728]
26. Bauermeister JA, Connochie D, Eaton L, Demers M, Stephenson R. Geospatial Indicators of Space and Place: A Review of Multilevel Studies of HIV Prevention and Care Outcomes Among Young Men Who Have Sex With Men in the United States. J Sex Res 2017;54(4-5):446-464 [FREE Full text] [doi: 10.1080/00224499.2016.1271862] [Medline: 28135857]
27. US Census Bureau, American Community Survey, 2016 American Community Survey 1-year estimates. URL: http://data.census.gov/cedsci [accessed 2019-02-02]
28. Agresti A. An introduction to categorical data analysis. Hoboken, New Jersey: WIley; 2018.
29. R Foundation. R: A language environment for statistical computing computer program. Vienna, Austria: R Foundation for Statistical Computing; 2013. URL: https://www.r-project.org/ [accessed 2020-05-20]
30. Venables W, Ripley B. Modern Applied Statistics with S. New York: Springer; 2002.
31. Noguchi K, Handley IM, Albarracín D. Participating in politics resembles physical activity: general action patterns in international archives, United States archives, and experiments. Psychol Sci 2011 Feb;22(2):235-242 [FREE Full text] [doi: 10.1177/0956797610393746] [Medline: 21177515]
32. Wallerstein N, Duran B. Community-Based Participatory Research Contributions to Intervention Research: The Intersection of Science and Practice to Improve Health Equity. Am J Public Health 2010 Apr;100(S1):S40-S46. [doi: 10.2105/ajph.2009.184036]

Abbreviations

IRR: incidence rate ratio
NLP: natural language processing
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by J Lortz, J Black; comments to author 10.01.20; revised version received 05.03.20; accepted 22.03.20; published 24.06.20.

Please cite as:
Stevens R, Bonett S, Bannon J, Chittamuru D, Slaff B, Browne SK, Huang S, Bauermeister JA
Association Between HIV-Related Tweets and HIV Incidence in the United States: Infodemiology Study
J Med Internet Res 2020;22(6):e17196
URL: https://www.jmir.org/2020/6/e17196
doi:10.2196/17196
PMID:32579119

Original Paper

User Perceptions of Different Electronic Cigarette Flavors on Social Media: Observational Study

Xinyi Lu[1], BSc; Long Chen[1], BA, BSc; Jianbo Yuan[1], PhD; Joyce Luo[2]; Jiebo Luo[1], PhD; Zidian Xie[3], PhD; Dongmei Li[3], PhD

[1]University of Rochester, Rochester, NY, United States
[2]Princeton University, Princeton, NJ, United States
[3]University of Rochester Medical Center, Rochester, NY, United States

Corresponding Author:
Dongmei Li, PhD
University of Rochester Medical Center
265 Crittenden Boulevard
CU 420708
Rochester, NY
United States
Phone: 1 5852767285
Email: Dongmei_Li@urmc.rochester.edu

Abstract

Background: The number of electronic cigarette (e-cigarette) users has been increasing rapidly in recent years, especially among youth and young adults. More e-cigarette products have become available, including e-liquids with various brands and flavors. Various e-liquid flavors have been frequently discussed by e-cigarette users on social media.

Objective: This study aimed to examine the longitudinal prevalence of mentions of electronic cigarette liquid (e-liquid) flavors and user perceptions on social media.

Methods: We applied a data-driven approach to analyze the trends and macro-level user sentiments of different e-cigarette flavors on social media. With data collected from web-based stores, e-liquid flavors were classified into categories in a flavor hierarchy based on their ingredients. The e-cigarette–related posts were collected from social media platforms, including Reddit and Twitter, using e-cigarette–related keywords. The temporal trend of mentions of e-liquid flavor categories was compiled using Reddit data from January 2013 to April 2019. Twitter data were analyzed using a sentiment analysis from May to August 2019 to explore the opinions of e-cigarette users toward each flavor category.

Results: More than 1000 e-liquid flavors were classified into 7 major flavor categories. The fruit and sweets categories were the 2 most frequently discussed e-liquid flavors on Reddit, contributing to approximately 58% and 15%, respectively, of all flavor-related posts. We showed that mentions of the fruit flavor category had a steady overall upward trend compared with other flavor categories that did not show much change over time. Results from the sentiment analysis demonstrated that most e-liquid flavor categories had significant positive sentiments, except for the beverage and tobacco categories.

Conclusions: The most updated information about the popular e-liquid flavors mentioned on social media was investigated, which showed that the prevalence of mentions of e-liquid flavors and user perceptions on social media were different. Fruit was the most frequently discussed flavor category on social media. Our study provides valuable information for future regulation of flavored e-cigarettes.

(J Med Internet Res 2020;22(6):e17280) doi:10.2196/17280

KEYWORDS

e-cigarette; flavor; social media

Introduction

Background

An electronic cigarette (e-cigarette) is a product for users to consume nicotine-related aerosols by heating a solution comprising propylene glycol or glycerol, nicotine, and flavoring chemicals [1]. The first e-cigarette product appeared in the market in 2003 and became popular in the United States in 2006-2007 [2]. E-cigarette products are commonly used as substitutes for conventional cigarettes. Consequently, conventional cigarette consumption in the United States has been decreasing in recent years [3]. The number of adult smokers declined from 20.9% in 2005 to 15.5% in 2016 [4]. Recently, people have tended to use e-cigarette products over conventional tobacco products, especially among adolescents and young adults [5]. According to the 2019 National Youth Tobacco Survey, among middle school students, the prevalence of e-cigarette use increased from 4.9% in 2018 to 10.5% in 2019, and for high school students, the number increased from 20.8% in 2018 to 27.5% in 2019 [6]. Approximately 72.2% of high school and 59.2% of middle school exclusive e-cigarette users used flavored e-cigarettes [6]. Flavored e-cigarette use among current young adult users (aged between 18 and 24 years) exceeds that of current older adult users (aged 25 years), and 81.5% of students aged between 12 and 17 years, used e-cigarettes because they came with attractive flavors [7]. Fruit, menthol or mint, candy, desserts, or other sweets were the most popular flavors among high school and middle school students [8].

Although perceived as safe to be ingested, flavorings in e-cigarettes remain to be a health concern, especially for adolescents, as they might not be safe to inhale. Different electronic cigarette liquid (e-liquid) flavors contain different flavoring chemicals [9]. Exposure to e-cigarette aerosols with flavorings could cause DNA damage, increased oxidative stress, inflammatory cytokine release, and epithelial barrier dysfunction [10,11]. A recent study showed that treating monocytic cells with different flavoring chemicals and flavored e-liquids without nicotine could cause different cytotoxicities [11]. E-cigarette flavorings could damage endothelial cells that line the interior of human blood vessels and may increase the risk of heart disease [12]. In particular, cinnamon and menthol flavors were more harmful than other flavors [11]. Diacetyl and 2,3-pentanedione, 2 of the most common flavoring ingredients in e-cigarettes, were found to have adverse effects on lung cells, which can impair lung function and cause a *popcorn lung* [9]. Another study suggested that the complexity of e-cigarettes on human health goes beyond respiratory and cardiac systems and may have significant implications on oral health [13].

Objective

Although most studies focus on the health effects of different e-liquid flavors, it has become very important to understand the relative proportions of mentions of different e-liquid flavors and user perceptions to reduce their potential health effects. Krüsemann et al [14] constructed an e-liquid flavor wheel based on flavor classifications using terms associated with e-cigarettes and flavors in research articles. Wang et al [15] investigated the e-cigarette content on Reddit without analyzing preferences and sentiments. A Twitter-related study focused on analyzing sentiments of general e-cigarette posts among different groups of users without considering e-cigarette flavors [16]. Currently, there have been few discussions about the e-liquid flavor classification, sentiment analysis, and trend analysis from the perspective of e-cigarette users on social media, which we explored in this study using Reddit and Twitter data.

With the emergence of new e-liquid flavors, the prevalence of mentions of e-liquid flavors might evolve over time, which needs to be monitored in a timely manner. In addition, it is important to understand how the perception of e-liquid flavors correlates with their prevalence. In this study, we employed a data-driven approach to characterize the dynamic changes in mentions of e-liquid flavors over time and to correlate them with the users' perceptions of each flavor category on social media. With the development of technology, social media have become popular in our daily lives. The number of American adults using social networking sites increased from 7% in 2005 to almost 65% in 2015 and is still increasing [17]. We collected data from Reddit, which is one of the most popular social media platforms [18]. Reddit posts usually contain more information because they have a much higher character limit compared with other social media platforms such as Twitter, thus allowing the discussions on Reddit to focus more on the opinions and experiences of users. Twitter posts have a lower character limit than Reddit posts. This advantage makes it easier to locate flavor-related sentences, which can lead to a more accurate sentiment analysis compared with using Reddit data. Therefore, we downloaded Twitter data using the public Twitter app programming interface to explore people's attitudes toward each flavor category. This study investigated the trends of mentions of e-liquid flavors and their perception on social media and provided some guidance for future research on the associations between the prevalence of e-cigarette flavors and user perceptions.

Methods

Electronic Cigarette Liquid Flavor Data Collection and Preprocessing

To obtain a full list of e-liquid flavor names, we searched the 3 major e-cigarette web-based stores, including *electrictobacconist.com*, *myvaporstore.com*, and *ecig-city.com*. The information on e-liquid flavors, including e-liquid brands, flavor names, and flavor ingredients, was manually recorded and curated. In addition to the unique flavor names (brand flavor names) from e-cigarette companies (eg, *Lava Flow* from *Naked 100* and *Green Goblin Salt* from *Oh My Gush*), specific flavor names (eg, mango, cotton candy, coffee) were also added to the list as people tend to mention specific flavor names more frequently on social media instead of using specific brand names. A total of 129 e-liquid brands and 1198 e-liquid flavors were collected.

Electronic Cigarette Liquid Flavor Classification

We applied a data-driven approach to construct a hierarchical e-liquid flavor classification, including 3 levels: major flavor categories (column 1 in Table 1), subcategories (column 2 in Table 1), and specific flavors (column 3 in Table 1). Specific flavor names were first standardized based on the flavor components and key ingredients because different brands or websites may use different names for the same flavor. As some specific e-cigarette flavor names mentioned in Reddit posts were sparse, the specific flavors were classified into several major categories. Similar to categories mentioned in the study by Krüsemann et al [14], 7 major flavor categories were generated: fruit, tobacco, menthol or mint, sweets, beverage, mixed, and others (Multimedia Appendix 1). Within each major flavor category, specific flavors were merged into several subcategories (middle level). For example, in the *fruit* flavor category, there are *tropical*, *berry*, *melon*, *mixed fruit*, and *others* flavor subcategories. In each subcategory, specific fruit flavors were listed. E-liquid flavors that contain more than one flavor ingredient in multiple categories were classified into the *mixed* category. All other flavors that could not be classified into previous categories were classified into the *others* flavor category.

Table 1. Classification of electronic cigarette liquid flavors.

Flavor categories and subcategories	List of specific flavors
Fruit	
Berry	Wildberry, currant, blackcurrant, blackberry, grape, raspberry, blueberry, strawberry, etc
Tropical	Mango, lychee, guava, passion fruit, pineapple, etc
Citrus	Grapefruit, lime, orange, lemon, etc
Melon	Cantaloupe, honeydew, melon, watermelon, etc
Mixed fruits	Mango apricot, apple melon, nana berry, etc
Others	Pomelo, papaya, apricot, dragon fruit, pomegranate, cucumber, kiwi, pear, cherry, peach, coconut, banana, apple, etc
Sweets	
Dessert	Mochi, pie, waffle, donut, mixed, cake, s'more, muffin, ice cream, cream, custard, macaron, granola, pastry, meringue, bread, cheesecake, cookie, etc
Candy	Lollypop, mixed, jelly bean, gummy bear, cotton candy, marshmallow, bubble gum, chocolate, etc
Others	Cereal, honey, caramel, etc
Beverage	
Coffee	Latte, mocha, cappuccino, espresso, coffee
Tea	Chai, tea, etc
Juice	Limeade, lemonade, apple juice, etc
Milk	Yogurt, milkshake, milk
Soft drinks	Cola, coke, soda, etc
Others	Energy drink, smoothie, etc
Tobacco	
Tobacco	Classic tobacco, Virginia tobacco, cigar, etc
Menthol or mint	
Menthol	Menthol
Mint	Mint, peppermint, spearmint
Mixed	
Mixed	Fruit + mint, fruit + sweets, fruit + beverage, fruit + beverage, sweets + other, sweets + mint, sweets + tobacco, fruit + tobacco, etc
Others	
Alcohol	Margarita, whiskey, rum, bourbon, cocktail, etc.
Nuts	Walnut, pecan, pistachio, hazelnut, almond, peanut butter, etc
Spice	Vanilla, cinnamon, etc
Others	Pure VG[a], pure PG[b], PG/VG, etc

[a] VG: vegetable glycerin.

[b] PG: propylene glycol.

Reddit

Reddit Data Collection and Preprocessing

Reddit posts from January 2013 to April 2019 were downloaded from *pushshift.io* (a website that publishes archive Reddit data) [19]. E-cigarette–related Reddit posts were obtained using keyword matching based on a list of e-cigarette–related keywords, including *e-cig*, *e-cigs*, *ecig*, *ecigs*, *electroniccigarette*, *ecigarette*, *ecigarettes*, *vape*, *vapers*, *vaping*, *vapes*, *e-liquid*, *ejuice*, *eliquid*, *e-juice*, *vapercon*, *vapeon*, *vapefam*, *vapenation*, and *juul* [20]. As a result, 2,865,467 e-cigarette–related Reddit posts were collected.

We further extracted e-liquid flavor–related Reddit posts from the e-cigarette dataset we collected. We applied the constructed flavor classification scheme and used the flavors and flavor names as the keywords to obtain posts that discuss e-liquid flavors. We observed that some flavor names were very likely to cause confusion and introduce noises such as *contact* and *punched*; thus, a denoise procedure was performed to filter out noisy flavor keywords [21]. For each confusing flavor name, sample posts were generated and manually labeled. The goal of the denoise procedure was to optimize the precision to 90% while maintaining an acceptable recall of ≥75%. If flavor names did not achieve the goal, they were excluded from further data analysis. After filtering out noisy flavor keywords, the final flavor subset contained 904,045 posts (Multimedia Appendix 2).

Temporal Analysis of Reddit

A temporal analysis was conducted to investigate the longitudinal trend of the intensity of discussion about each e-cigarette flavor category using the monthly post counts from January 2013 to April 2019. To further investigate competitions and substitutions between flavors, data were further normalized by the total number of flavor subset post counts in each month.

Twitter

Twitter Data Collection and Preprocessing

Using the same e-cigarette–related keywords, Twitter streaming data were downloaded from May 31 to August 22, 2019, leading to a collection of 2,757,860 e-cigarette–related Twitter posts.

To obtain e-cigarette flavor–related feedback from e-cigarette users instead of e-cigarette promotions from official accounts, promotion-related Twitter posts were filtered out in 2 steps. First, Twitter IDs that contain promotion-related keywords were eliminated, including *dealer*, *deal*, *store*, *supply*, *e-cig*, *e-cigs*, *ecig*, *ecigs*, *electroniccigarette*, *ecigarette*, *ecigarettes*, *vape*, *vapers*, *vaping*, *vapes*, *e-liquid*, *ejuice*, *eliquid*, *e-juice*, *vapercon*, *vapeon*, *vapefam*, *vapenation*, and *juul*. Second, Twitter posts that contain promotion-like keywords (eg, customer, promotion, discount, sale, and free shipping) were eliminated. After these 2 filtering procedures, a subset of 2,530,048 e-cigarette–related posts remained. To obtain an accurate result on the Twitter sentiment analysis of each flavor category, another flavor-related filtering was conducted by using keywords from the previous e-liquid flavor list Multimedia Appendix 1). As a result, we obtained an e-liquid flavor–related Twitter subset that contains 21,389 posts, and each post contains only 1 flavor keyword. The complete procedures of Twitter data preprocessing are shown in Multimedia Appendix 2.

Sentiment Analysis of Electronic Cigarette Liquid Flavors on Twitter

A sentiment analysis is a contextual analysis of sentences and paragraphs, which can extract subjective attitudes and opinions from the source material. The Valence Aware Dictionary and sEntiment Reasoner was used as the sentiment analyzer to extract the thoughts and opinions of e-cigarette users on each flavor category from Twitter posts [22]. We obtained a sentiment score for each tweet, and the average sentiment score was calculated for each flavor category. On the basis of the suggested threshold for determining the sentiment (positive, neutral, and negative), tweets with sentiment scores in the range of −1.00 to −0.05 were classified as negative posts, tweets with sentiment scores in the range of −0.05 to +0.05 (not include −0.05 and +0.05) were classified as neutral posts, and tweets with sentiment scores in the range of +0.05 to +1.00 were classified as positive posts [22]. The number of posts for each flavor category with positive, neutral, and negative sentiments was normalized by the total number of posts for each flavor category to explore the distribution of the sentiment results for each flavor category. In addition, comparisons between the proportions of positive and negative Twitter posts within each flavor category were conducted using 2 proportion z tests in SAS version 9.4 (SAS Institute Inc) to determine the significant differences between the proportion of positive sentiments and the proportion of negative sentiments. All tests were two sided, with a significance level of 5%. The original *P* values were adjusted using the Bonferroni method to account for multiplicity.

Results

Prevalence of Electronic Cigarette Flavors Mentioned on Reddit

E-liquid flavors were classified into different flavor categories. Table 1 provides an overview of the multiple levels of classifications, including major flavor categories, subcategories, and specific flavors. The proportions of hierarchical 3-layer flavor classifications are shown in Multimedia Appendix 1. Using the filtered Reddit data, Table 2 shows the percentage distribution of major flavor categories as well as detailed percentage distributions of 4 major flavor subcategories. As shown in Table 2, the *fruit* and *sweets* categories were the most mentioned and popular flavor categories on Reddit, followed by *beverage*, *menthol or mint*, *tobacco*, *others*, and *mixed*. In addition, *berry* was the dominant subcategory with a percentage of 45.84% (7,27,404/15,86,926) over all fruit-related subcategories; *dessert* was dominant in the *sweets* subcategory with a percentage of 47.43% (1,89,946/400,500). *Coffee* (118,129/2,77,005, 42.65%) and *tea* (1,07,449/2,77,005, 38.79%) were the most popular subcategories in the *beverage* flavor category. In the *menthol or mint* flavor category, the *menthol* (1,73,641/2,29,817, 75.56%) subcategory was more popular than the *mint* (56,176/2,29,817, 24.44%) subcategory. The percentage of post count for specific flavors in Multimedia Appendix 1 showed that the *strawberry* flavor was the most mentioned in the *fruit* category, and the *vanilla* flavor had the most mentions in the *sweets* category.

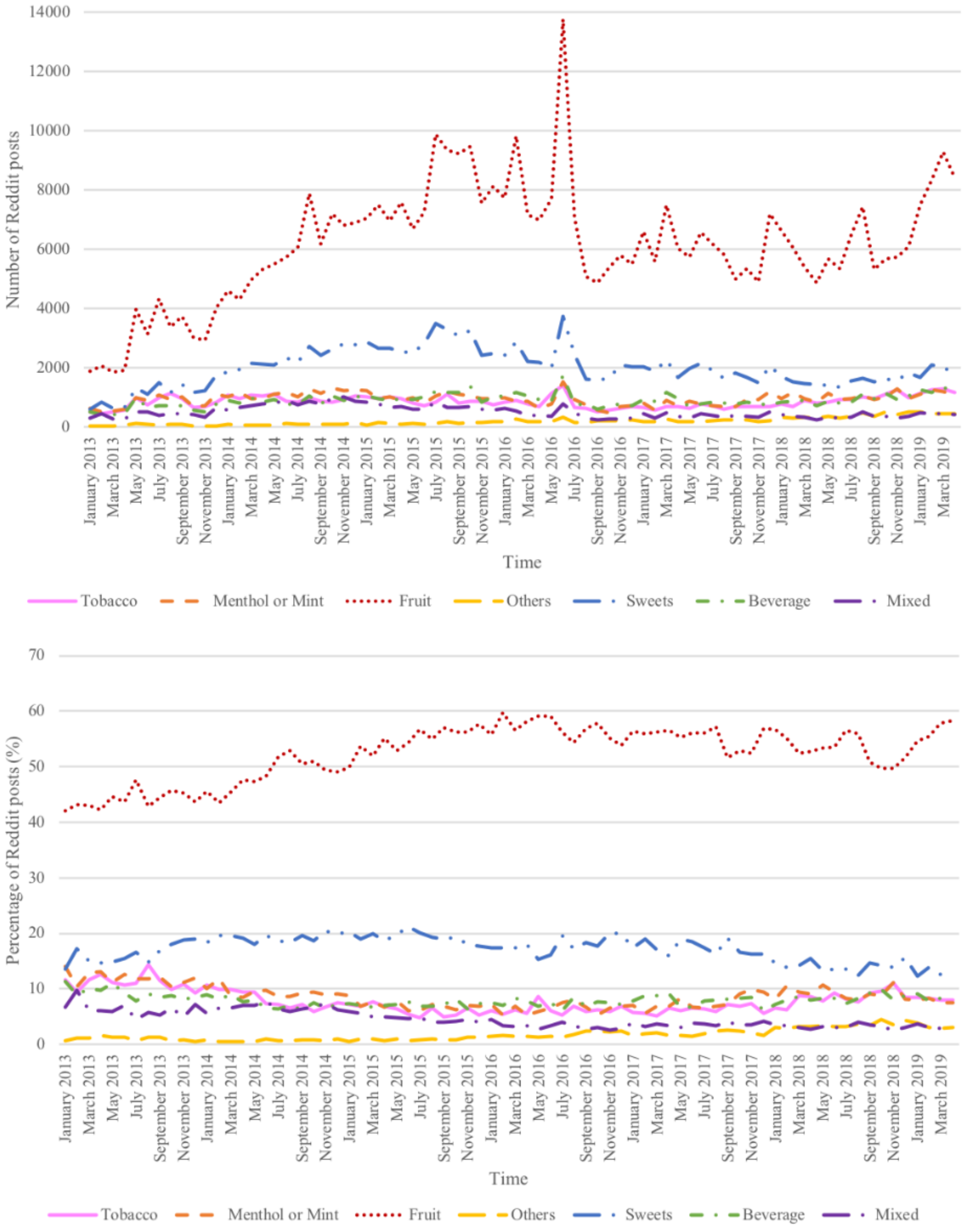
Table 2. Percentage distribution of e-liquid flavors mentioned on Reddit.

Flavor categories and subcategories	Count, n (%)
Fruit	1,586,926 (58.15)
Berry	727,404 (45.84)
Others	374,886 (23.62)
Tropical	190,639 (12.01)
Melon	138,876 (8.75)
Mixed fruit	102,618 (6.47)
Citrus	52,503 (3.31)
Sweets	400,500 (14.67)
Dessert	189,946 (47.43)
Others	113,688 (28.39)
Candy	96,866 (24.18)
Beverage	277,005 (10.15)
Coffee	118,129 (42.64)
Tea	107,449 (38.80)
Milk	21,764 (7.85)
Juice	21,117 (7.62)
Soft drink	5377 (1.94)
Others	3169 (1.14)
Menthol or mint	229,817 (8.42)
Menthol	173,641 (75.56)
Mint	56,176 (24.44)
Tobacco	163,377 (5.99)
Others	43,922 (1.61)
Mixed	24,769 (0.91)

Temporal Analysis of Mentions of Electronic Cigarette Liquid Flavors on Reddit

With the availability of new e-liquid flavors and the changes in the perception of users on e-liquid flavors over time, it is important to examine the temporal changes of e-liquid flavors mentioned on social media, so that the appropriate regulation of e-liquid flavors could be applied. To address this, we used the Reddit data collected from January 2013 to April 2019. By examining the number of Reddit posts on each e-liquid flavor, we observed a relatively increasing trend on the *fruit* category, whereas it remained relatively constant for the other flavor categories (Figure 1). The temporal analysis also showed a significant spike for each flavor category in June 2016 (Figure 1). This spike coincided with the US Food and Drug Administration (FDA), extending its regulatory authority to all tobacco products, including e-cigarettes and pipe tobacco on June 16, 2016. The number of posts mentioning e-liquid flavors on Reddit after June 2016 was much less than that before June 2016 but still showed an upward trend. After normalization to all posts mentioning flavors, the *fruit* flavor category had a steady upward trend from January 2013 to June 2016 and then showed some fluctuation. Most of the other flavor categories did not show significant trends, with *sweets* and *mixed* categories having a relatively noticeable decreasing trend.

Figure 1. The longitudinal trend of e-liquid flavors mentioned on Reddit, including the number of Reddit posts and the proportion of Reddit posts of each flavor category from January 2013 to April 2019. The proportion is normalized by the monthly total flavor-related post count. e-liquid: electronic cigarette liquid.



Sentiment Analysis of Electronic Cigarette Liquid Flavors Mentioned on Twitter

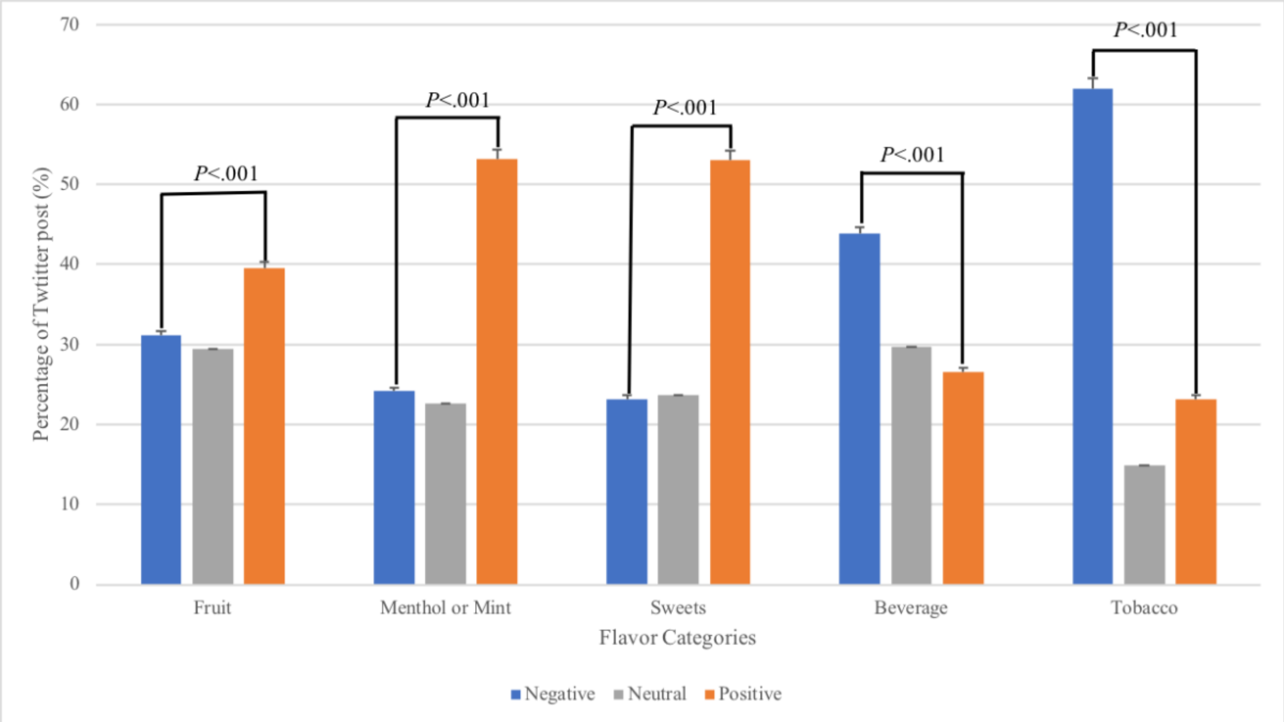
Although there are more than 7000 e-liquid flavors available on the market, it is important to examine the perception of public users to different e-liquid flavors, which will help us understand their prevalence on social media. As the character limit of Reddit subreddits is 40,000, it was difficult to locate the sentences only about e-liquid flavors in long Reddit messages. In this case, sentiment analysis precision for e-liquid flavors on Reddit cannot be ensured. Thus, Twitter data with a character limit of 280 were used instead for the sentiment analysis for more precise results. To obtain the most recent sentiments toward e-liquid flavors, we decided to use Twitter data from May 31 to August 22, 2019. As the total post counts in the *others* (276 posts) and the *mixed* (126 posts) flavor categories are much lower than other flavor categories and because e-cigarette flavors in these 2 categories were miscellaneous and diverse, we excluded these 2 categories in the sentiment analysis. The sentiment analysis on each e-liquid flavor category using Twitter data showed that, on average, the posts with the *sweets*, *menthol or mint*, and *fruit* flavor categories had positive sentiments, whereas the posts with the *tobacco* and *beverage* flavor categories showed negative sentiments (Table 3).

On the basis of thresholds of +/- 0.05 for positive and negative sentiments, each post was classified into positive, negative, or neutral sentiments. Within each flavor category, the proportion of posts with different sentiments was calculated (Figure 2). The *fruit*, *menthol or mint*, and *sweets* flavor categories had a significantly higher proportion of positive posts than negative posts (*P*<.001). In contrast, the *beverage* and *tobacco* flavor categories had a significantly higher proportion of negative posts than positive posts (*P*<.001).

Table 3. Sentiment analysis of flavor categories on Twitter.

Flavor category	Total post count, N	Average sentiment score, mean (SD)
Fruit	9852	0.074 (0.453)
Menthol or mint	4582	0.128 (0.379)
Sweets	3190	0.156 (0.442)
Beverage	1805	−0.090 (0.451)
Tobacco	1555	−0.134 (0.440)

Figure 2. Sentiment analysis of the flavor categories mentioned on Twitter. Error bars represent the estimated SDs.



Discussion

Principal Findings

In this study, by mining e-cigarette–related posts on social media (Reddit and Twitter), we showed that *fruit* and *sweets* flavors were the most popular flavors mentioned on social media, which remained to be the top popular flavors mentioned on social media over time. In addition, the sentiment analysis result showed significant positive attitudes toward *fruit* and *sweets* flavor categories. Although there were significant negative attitudes toward *tobacco* and *beverage* flavor categories, these 2 flavors were less mentioned on social media. Therefore, the perception of e-cigarette flavors on social media positively correlates with their popularity, that is, a positive sentiment correlates with high popularity on social media.

Our temporal studies showed that *fruit* and *sweets* have constantly been the most frequently mentioned e-liquid flavors on social media from 2013 to 2019, suggesting that the prevalence of e-liquid flavors mentioned on social media has not changed much over time. However, with the implementation of the e-cigarette flavor enforcement policy by the FDA on February 6, 2020, which bans all cartridge-based–flavored e-cigarettes except tobacco and menthol flavors, the popularity of e-liquid flavors will change accordingly. In addition, with the availability of other flavored e-cigarette products (such as disposable e-cigarette devices) on the market after the FDA flavor enforcement policy, it is hard to predict which flavors will become popular, which warrants further investigation.

In this study, we provided a way to timely monitor the prevalence of e-liquid flavors mentioned on social media as well as to identify the positive correlation between the perception of e-liquid flavors and their prevalence. With the power of this surveillance system, current regulation on flavored e-cigarettes could be modified or updated in a timely manner, and more importantly, the epidemic of e-cigarette use, especially among the youth, could be ameliorated to protect public health.

Comparison With Prior Work

Compared with the previous flavor wheel that was based on manually reviewed papers, the proposed e-liquid flavor classification in this study has better coverage of the up-to-date e-cigarette flavors with practical frequency distributions because it was constructed based on products on the web and discussions on social media [14]. As a result, we revised the flavor classification in several ways. First, the major flavor categories were merged from 13 into 7 categories [14]. For example, we combined the *candy*, *dessert*, and *other sweets* categories into the other sweets category of *sweets* as most of these flavors have sweetener ingredients. In addition, subcategories were created to form a more detailed flavor classification. For example, within the *fruit* flavor category, besides the existing second-level subcategories *tropical*, *berry*, *citrus*, and *others*, a *melon* subcategory was added because of the frequent mentions of melon-related flavors on social media. Together, the revised flavor classification contains more complete and the most up-to-date e-cigarette flavors.

A previous study showed the prevalence of the mentions of e-liquid flavor use on Reddit [15]. Instead of showing the yearly nominal post counts of each flavor category, our temporal analysis used the total flavor-related Reddit post counts to normalize the monthly data to see the trend of mentions of e-liquid flavors. Similar to previous findings [15], our results showed that the *fruit* flavor category was the dominant flavor on Reddit, followed by *sweets*. Another study showed that the *fruit* flavor category was mentioned most frequently on Facebook, followed by *sweet* and *cream* (classified as the *sweets* flavor category in our study) [23]. Compared with the findings from the previous study [15], the *beverage* flavor had a slightly larger portion of mentions recently. The percentage distribution of the *beverage* flavor increased from 7% to 10% from 2015 to 2019. Although this increase is small, which could be because of random error, there might be 2 other reasons. First, it could be caused by increased discussions about the *beverage* flavors on Reddit. Second, our *beverage* flavor-filtering keyword list contains more flavors. Besides the *coffee* and *tea* flavors, we added the *juice*, *soft drinks*, and *energy drinks* flavor subcategories in the *beverage* category, which might cause the *beverage* flavor category to have a larger percentage of mentions on Reddit [15].

On June 16, 2016, the FDA finalized a rule that extended its regulatory authority to all tobacco products, including e-cigarettes and pipe tobacco, as part of its goal to improve public health. This new rule also restricted youth access to newly regulated tobacco products [24]. Due to this new rule released by the FDA, more flavor-related e-cigarette Reddit posts were mentioned during that month, causing a significant spike in Figure 1. These data also demonstrated that social media data could accurately and timely reflect what happens in real life. There was a slightly increasing trend in the *fruit* category in Figure 1, which suggests that people tend to talk about fruit-related e-liquid flavors more frequently on Reddit compared with other flavor categories. A recent study using the Population Assessment of Tobacco and Health Wave 3 data showed that 52.8% of youths used fruit-flavored electronic nicotine delivery systems, which is consistent with our findings on the prevalence of fruit flavors (58%) mentioned using Reddit data [25]. Our sentiment analysis showed positive attitudes toward the fruit flavor category. Thus, our data suggest that the regulation of e-cigarette use should focus on the restriction of e-cigarette use with fruit flavors.

There have been few studies on flavor sentiment analysis on social media [16,26]. One study focused on e-cigarette–related Twitter post sentiments among users with respect to different genders and age ranges [16]. However, flavor-related sentiments were not discussed. Another study showed the sentiments on e-liquid flavors mentioned on JuiceDB, but only focusing on the 2 most popular flavors, *fruit* and *sweet* flavors [26]. Our results were similar to the sentiment analysis from JuiceDB, showing that the overall sentiments for those 2 flavors were positive. More importantly, our results performed a sentiment analysis on other flavor categories, thus providing a more comprehensive picture of the perception of e-liquid flavors on social media.

Limitations

In this study, Reddit and Twitter data were used to answer different questions considering their characteristics. A Reddit post has a relatively large character limit, which makes it an ideal data source to examine the experience users have with e-cigarettes. In addition, Reddit posts were available from January 2013 to April 2019, which made it possible to examine the temporal changes in popular e-liquid flavors. However, as Reddit posts have a 40,000-character limit, it is difficult to obtain accurate sentiment results for specific e-liquid flavors. Twitter posts have a much smaller character limit (280 characters) compared with Reddit, which makes it ideal for a sentiment analysis. Switching social media platforms from Reddit to Twitter in the process of a sentiment analysis may create some bias. The attitudes of users toward e-cigarette flavors on Twitter might be slightly different from those on Reddit.

Although there are more than 7000 e-liquid flavors on the market [27], just over 1000 specific e-liquid flavors were included in our dataset. In the future, more e-liquid brands and flavors will be collected for a more complete dataset, which can generate more detailed flavor classification. More analyses could be conducted by using additional social media platforms, such as Facebook and Instagram.

User demographic information, including gender, age, and ethnicity, is not available from social media data, which might limit further analysis of the prevalence of e-liquid flavors mentions among different demographic groups.

In general, the information from social media data is noisy, which could bias the results. However, our results are consistent with the results from the national survey data, which indicate the reliability of our conclusions.

With the announcement of the e-cigarette flavor enforcement policy by the FDA at the beginning of 2020, the prevalence of e-cigarette flavors will likely evolve as most flavors (such as fruit and sweets) are banned and as disposable e-cigarette devices with different flavors become available on the market. Understanding which flavors are the most popular after the flavor enforcement policy will be very important, which awaits further investigation.

Conclusions

Although the popularity of e-liquid flavors has been reported in previous studies, this study showed the longitudinal changes in the prevalence of mentions of e-liquid flavors as well as the perceptions users have of different e-liquid flavors. The findings from this study will provide the most updated information about the popular e-liquid flavors mentioned on social media and how the perceptions of different e-liquid flavors are correlated with their prevalence on social media, which could guide further regulation of e-cigarette flavors. Our results showed that the prevalence and perceptions of different flavors on social media were different. The findings of this study have several valuable applications. Social media data can be used to inform researchers or policymakers about the prevalence of mentions of different e-liquid flavors in a timely manner and provide some guidance on the regulation of flavored e-cigarettes. Although the FDA has restricted the sale of unauthorized flavored cartridge-based e-cigarettes, flavored e-cigarettes in other formats still exist in the market, such as disposable e-cigarette devices. Here, we provided a cost-effective surveillance system to monitor the prevalence of e-liquid flavors over time.

Acknowledgments

This study was supported by the National Cancer Institute of the National Institutes of Health and the FDA Center for Tobacco Products under award number U54CA228110.

Authors' Contributions

XL, ZX, and DL conceived and designed the study. XL and LC analyzed the data. XL wrote the manuscript. JY, JL, JL, ZX, and DL assisted with the interpretation of the analyses and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of Data Pre-processing Procedures.
[PNG File , 117 KB - jmir_v22i6e17280_app1.png]

Multimedia Appendix 2
E-liquid Flavors mentioned on Reddit.
[DOCX File , 20 KB - jmir_v22i6e17280_app2.docx]

References

1. Grana R, Benowitz N, Glantz SA. E-cigarettes: a scientific review. Circulation 2014 May 13;129(19):1972-1986 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.007667] [Medline: 24821826]
2. Consumer Advocates for Smoke-Free Alternatives Association. 2018. A Historical Timeline of Electronic Cigarettes URL: http://www.casaa.org/historical-timeline-of-electronic-cigarettes/ [accessed 2020-05-05]
3. Hoffman SJ, Mammone J, van Katwyk SR, Sritharan L, Tran M, Al-Khateeb S, et al. Cigarette consumption estimates for 71 countries from 1970 to 2015: systematic collection of comparable data to facilitate quasi-experimental evaluations of national and global tobacco control interventions. Br Med J 2019 Jun 19;365:l2231 [FREE Full text] [doi: 10.1136/bmj.l2231] [Medline: 31217224]
4. Centers for Disease Control and Prevention. 2018. Smoking is down, but almost 38 million American adults still smoke URL: https://www.cdc.gov/media/releases/2018/p0118-smoking-rates-declining.html [accessed 2020-05-05]
5. Levy D, Yuan Z, Li Y. The prevalence and characteristics of e-cigarette users in the US. Int J Environ Res Public Health 2017 Oct 11;14(10):1200 [FREE Full text] [doi: 10.3390/ijerph14101200] [Medline: 29019917]
6. Gentzke AS, Creamer M, Cullen KA, Ambrose BK, Willis G, Jamal A, et al. Vital signs: tobacco product use among middle and high school students-United States, 2011-2018. MMWR Morb Mortal Wkly Rep 2019 Feb 15;68(6):157-164 [FREE Full text] [doi: 10.15585/mmwr.mm6806e1] [Medline: 30763302]
7. Murthy VH. E-cigarette use among youth and young adults: a major public health concern. JAMA Pediatr 2017 Mar 1;171(3):209-210. [doi: 10.1001/jamapediatrics.2016.4662] [Medline: 27928577]

8. Cullen KA, Gentzke AS, Sawdey MD, Chang JT, Anic GM, Wang TW, et al. E-cigarette use among youth in the United States, 2019. J Am Med Assoc 2019 Nov 5:- epub ahead of print. [doi: 10.1001/jama.2019.18387] [Medline: 31688912]
9. Allen JG, Flanigan SS, LeBlanc M, Vallarino J, MacNaughton P, Stewart JH, et al. Flavoring chemicals in e-cigarettes: diacetyl, 2,3-pentanedione, and acetoin in a sample of 51 products, including fruit-, candy-, and cocktail-flavored e-cigarettes. Environ Health Perspect 2016 Jun;124(6):733-739 [FREE Full text] [doi: 10.1289/ehp.1510185] [Medline: 26642857]
10. Gerloff J, Sundar IK, Freter R, Sekera ER, Friedman AE, Robinson R, et al. Inflammatory response and barrier dysfunction by different e-cigarette flavoring chemicals identified by gas chromatography-mass spectrometry in e-liquids and e-vapors on human lung epithelial cells and fibroblasts. Appl In Vitro Toxicol 2017 Mar 1;3(1):28-40 [FREE Full text] [doi: 10.1089/aivt.2016.0030] [Medline: 28337465]
11. Muthumalage T, Prinz M, Ansah KO, Gerloff J, Sundar IK, Rahman I. Inflammatory and oxidative responses induced by exposure to commonly used e-cigarette flavoring chemicals and flavored e-liquids without nicotine. Front Physiol 2017;8:1130 [FREE Full text] [doi: 10.3389/fphys.2017.01130] [Medline: 29375399]
12. Lee WH, Ong S, Zhou Y, Tian L, Bae HR, Baker N, et al. Modeling cardiovascular risks of e-cigarettes with human-induced pluripotent stem cell-derived endothelial cells. J Am Coll Cardiol 2019 Jun 4;73(21):2722-2737. [doi: 10.1016/j.jacc.2019.03.476] [Medline: 31146818]
13. Kim SA, Smith S, Beauchamp C, Song Y, Chiang M, Giuseppetti A, et al. Cariogenic potential of sweet flavors in electronic-cigarette liquids. PLoS One 2018;13(9):e0203717 [FREE Full text] [doi: 10.1371/journal.pone.0203717] [Medline: 30192874]
14. Krüsemann EJ, Boesveldt S, de Graaf K, Talhout R. An e-liquid flavor wheel: a shared vocabulary based on systematically reviewing e-liquid flavor classifications in literature. Nicotine Tob Res 2019 Sep 19;21(10):1310-1319 [FREE Full text] [doi: 10.1093/ntr/nty101] [Medline: 29788484]
15. Wang L, Zhan Y, Li Q, Zeng D, Leischow S, Okamoto J. An examination of electronic cigarette content on social media: analysis of e-cigarette flavor content on Reddit. Int J Environ Res Public Health 2015 Nov 20;12(11):14916-14935 [FREE Full text] [doi: 10.3390/ijerph121114916] [Medline: 26610541]
16. Resende EC, Culotta A. Computer Science: Illinois Institute of Technology. 2015. A Demographic and Sentiment Analysis of E-cigarette Messages on Twitter URL: http://cs.iit.edu/~culotta/pubs/resende15demographic.pdf [accessed 2020-05-05]
17. Perrin A. Pew Research Center. 2015. Social Media Usage: 2005-2015 URL: https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/ [accessed 2020-05-05]
18. Shearer E, Matsa K. Pew Research Center. 2018. News Use Across Social Media Platforms URL: https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/ [accessed 2020-05-05]
19. Baumgarten J. Reddit-PushShift. Directory Contents URL: https://files.pushshift.io/reddit/ [accessed 2019-11-04]
20. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for coding tobacco-related twitter data: a systematic review. J Med Internet Res 2017 Mar 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]
21. Stieglitz S, Mirbabaie M, Ross B, Neuberger C. Social media analytics–challenges in topic discovery, data collection, and data preparation. Int J Inform Manage 2018 Apr;39:156-168. [doi: 10.1016/j.ijinfomgt.2017.12.002]
22. Hutto CJ, Gilbert E.. Vader: a Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; June 1-4, 2014; Ann Arbor, Michigan, USA p. A.
23. Zhou J, Zhang Q, Zeng DD, Tsui KL. Influence of flavors on the propagation of e-cigarette-related information: social media study. JMIR Public Health Surveill 2018 Mar 23;4(1):e27 [FREE Full text] [doi: 10.2196/publichealth.7998] [Medline: 29572202]
24. US Food and Drug Administration. 2016. The Facts on the FDA's New Tobacco Rule URL: https://www.fda.gov/consumers/consumer-updates/facts-fdas-new-tobacco-rule [accessed 2020-05-05]
25. Schneller LM, Bansal-Travers M, Goniewicz ML, McIntosh S, Ossip D, O'Connor RJ. Use of flavored e-cigarettes and the type of e-cigarette devices used among adults and youth in the US-results from wave 3 of the population assessment of tobacco and health study (2015-2016). Int J Environ Res Public Health 2019 Aug 20;16(16):2991 [FREE Full text] [doi: 10.3390/ijerph16162991] [Medline: 31434229]
26. Li Q, Wang C, Liu R, Wang L, Zeng DD, Leischow SJ. Understanding users' vaping experiences from social media: initial study using sentiment opinion summarization techniques. J Med Internet Res 2018 Aug 15;20(8):e252 [FREE Full text] [doi: 10.2196/jmir.9373] [Medline: 30111530]
27. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul;23(Suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]

Abbreviations

e-cigarette: electronic cigarette
e-liquid: electronic cigarette liquid
FDA: Food and Drug Administration

Edited by G Eysenbach; submitted 02.12.19; peer-reviewed by E O'Brien, L Sbaffi; comments to author 20.01.20; revised version received 15.03.20; accepted 29.03.20; published 24.06.20.

Please cite as:
Lu X, Chen L, Yuan J, Luo J, Luo J, Xie Z, Li D
User Perceptions of Different Electronic Cigarette Flavors on Social Media: Observational Study
J Med Internet Res 2020;22(6):e17280
URL: http://www.jmir.org/2020/6/e17280/
doi:10.2196/17280
PMID:32579123

Original Paper

Exploring Public Awareness of Overwork Prevention With Big Data From Google Trends: Retrospective Analysis

Ro-Ting Lin[1], PhD; Yawen Cheng[2], ScD; Yan-Cheng Jiang[1]

[1]Department of Occupational Safety and Health, College of Public Health, China Medical University, Taichung, Taiwan

[2]Institute of Health Policy and Management, College of Public Health, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Ro-Ting Lin, PhD
Department of Occupational Safety and Health
College of Public Health, China Medical University
No 91, Hsueh-Shih Road
Taichung, 40402
Taiwan
Phone: 886 4 2205 3366 ext 6215
Email: roting@ntu.edu.tw

Abstract

Background: To improve working conditions and prevent illness and deaths related to overwork, the Taiwanese government in 2015, 2016, and 2018 amended regulations regarding working time, overtime, shifts, and rest days. Such policy changes may lead to a rising public awareness of overwork-related issues, which may in turn reinforce policy development.

Objective: This study aimed to investigate to what extent public awareness of overwork-related issues correlated with policy changes.

Methods: Policies, laws, and regulations promulgated or amended in Taiwan between January 2004 and November 2019 were identified. We defined 3 working conditions (overwork, long working hours, and high job stress) related to overwork prevention, generated a keyword for each condition, and extracted the search volumes for each keyword on the Google search engine as proxy indicators of public awareness. We then calculated the monthly percentage change in the search volumes using the Joinpoint Regression Program.

Results: Apparent peaks in search volumes were observed immediately after policy changes. Especially, policy changes in 2010 were followed by a remarkable peak in search volumes for both overwork and working hours, with the search volumes for overwork increased by 29% per month from June 2010 to March 2011. This increase was preceded by the implementation of new overwork recognition guidelines and media reports of several suspected overwork-related events. The search volumes for working hours also steadily increased, by 2% per month in September 2013 and afterward, reaching a peak in January 2017. The peak was likely due to the amendment to the Labor Standards Act, which called for “1 fixed and 1 flexible day off per week,” in 2016. The search volumes for job stress significantly increased (P=.026) but only by 0.4% per month since March 2013.

Conclusions: Over the past 15 years, Taiwanese authorities have revised and implemented several policies to prevent overwork-related health problems. Our study suggests a relationship between the implementation of policies that clearly defined the criteria for overwork and working hours and the rising public awareness of the importance of overwork prevention and shorter working hours.

(J Med Internet Res 2020;22(6):e18181) doi:10.2196/18181

KEYWORDS

overwork; working hours; policy; big data

Introduction

Overwork—including long hours or high physical or psychological effort or strain—may lead to the occurrence of cerebrovascular and cardiovascular diseases, mental disorders, and even *karoshi* (occupational sudden death) [1-3]. The first *karoshi* case was reported in Japan in 1969: the death of a 29-year-old male worker in a shipping department due to a stroke [4]. An increasing number of overwork-related events have been reported ever since in Japan, South Korea, and

Taiwan, where the culture of long working hours is prevalent and even valued [5].

In Taiwan, several apparent overwork-related cases were reported in the period between 2009 and 2011 but their workers' compensation claims were disapproved by the Bureau of Labor Insurance. The ways the labor insurance authorities handled such disputes ignited public criticisms. Finally, these workers' family members turned to legislators for help, pushing the governments to revise their policies and regulations concerning overwork [6]. In response to societal concerns, the Taiwanese government has modified both the policies for workers' compensation benefits and the prevention policies for overwork [6]. With regard to workers' compensation, the government has loosened the recognition guidelines since the late 2010 to allow more cases to be compensated by the labor insurance system. With regard to prevention of overwork problems, the government adopted new prevention measures in the Occupational Safety and Health Act in 2013; revised legal regular working hour limit stipulated in the Labor Standards Act in 2015 from 84 hours over 2 weeks to 40 hours per week; and twice—in 2016 and again in 2018—amended regulations pertaining to regular working time, overtime, annual leave, shifts, and rest days. Through these actions, the government has demonstrated its commitment to preventing overwork by shortening working hours and improving working conditions.

In addition to regulatory reforms, for the prevention of overwork problems, it is important to enhance public awareness of health risks associated with overwork and other adverse working conditions, such as long working hours and high job stress [7,8]. On the one hand, policy changes may be the results of increased social concerns. On the other hand, the development of policies could further enhance social awareness. According to the knowledge, attitude, and practice theory [9], it can be expected that higher social awareness will lead to the acquisition of information and knowledge and the cultivation of better attitudes toward preventative practices. For example, clear regulation of a company's working hours could facilitate better understanding of how to prevent overwork among its employees. However, public awareness levels are not easily assessed.

The use of internet and online search engines such as Google, Yahoo!, and Bing has become one of the most popular means to obtain knowledge on specific topics, which further enhance people's attitudes, behaviors, and decisions [10]. A previous study has identified that in 2012, around 55% of internet users in the United States used search engines to obtain health-related information [11]. Among all search engines, Google is the most popular in many countries, and in Taiwan about 94% of internet users use Google as their primary search engine [12]. The Google Trends website reports on recorded Google search volumes, and has been widely used to measure public awareness and interest in specific issues, such as cancer screening and crude oil prices [13,14], to assess the impact of celebrity on public behavior [15], and to conduct health assessment studies [16]. Thus, Google search volumes extracted from the Google Trends platform can be a proxy indicator of public awareness of overwork-related issues.

Previous studies in Japan and South Korea have mentioned that public awareness of overwork is likely to be enhanced both by reports of the occurrence of overwork-related diseases and by recent changes in overwork-related policy [7,17]. While this assumption has been proposed, no study has yet demonstrated whether or not public awareness has been influenced by policy changes and reported overwork incidents. We hypothesized that public awareness increased following changes in policies which clearly address overwork-related risk factors. We therefore aimed to use Google search volumes to investigate whether and to what extent the public's awareness of overwork prevention has been influenced by policy changes and the occurrence of overwork-related events.

Methods

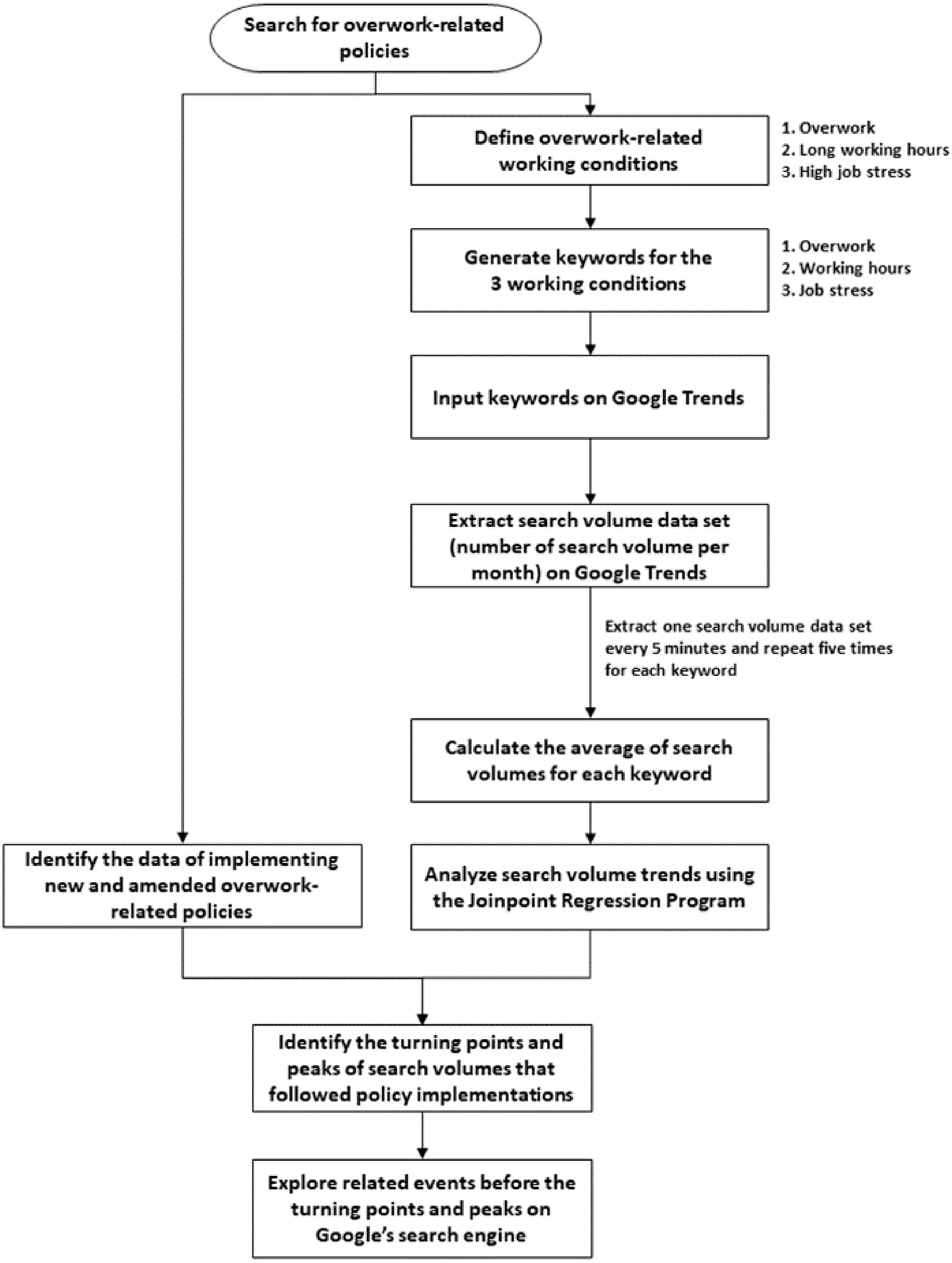
Study Process

Figure 1 shows our study process. We searched for policies, laws, and regulations promulgated or amended in Taiwan during the period between January 2004 and November 2019. We then identified 2 working conditions related to the aforementioned policy changes: (1) overwork, which has no official definition in Taiwan but is understood to describe both working too long (too many hours) and as working too hard (physically or psychologically); and (2) long working hours, which are regulated through both regular working time and overtime policies. We further identified a third working condition related to the 2004–2019 policy changes but had no direct quantitative measure: (3) high job stress. We treated overwork and long working hours as the main topics of interest, and high job stress as a reference topic (relevant to overwork but not addressed in the amendments of Labor Standards Act). To measure these 3 working conditions, we identified a keyword for each condition.

We used the keywords to extract search volumes from the Google Trends platform [18], which is used to record big data related to search volumes. The data on search volumes were scaled on a range of 0-100, with the scaled number shown as relative proportional to all searches. The website has 4 filter options: region, time, category, and search type; for these filter options, we selected Taiwan, January 2004–November 2019, all categories, and web search, respectively.

We noticed the variability of search volumes while we repeatedly searched on the Google Trends platform. The Google Trends sampled data on searches correspond to billions of Google searches per day. However, the algorithms behind the platform were not available to the general public [19]. We, therefore, developed our search strategy, repeated data extraction, and smoothed data by taking averages. We extracted one search volume data set every 5 minutes and repeated five times for each keyword. We then calculated the average of search volumes for each keyword and analyzed search volume trends using the Joinpoint Regression Program [20]. Next, we identified the turning points and peaks of search volumes that followed policy implementations. For points or peaks that did not follow policy changes, we used Google's search engine to explore related events before the turning points and peaks.

Figure 1. The study process.



Study Period

Our study period was from January 2004 to November 2019. January 2004 was the earliest data we could access from the Google Trends platform, and 2004 was also the year when Taiwan's Recognition Guidelines for Overwork-Related Diseases were first amended.

Data Analysis

First, we calculated the average monthly search volume for each keyword. Second, we used the Joinpoint Regression Program to estimate the web search volume trends over time [20]. The Joinpoint Regression Program allows us to analyze trend data and test joinpoints (or apparent changes or peaks) across the whole study period. Data captured as 0 were adjusted to 0.5 to fit all data on a log-linear Poisson regression model. Bayesian information criterion was used to identify turning points and select the best model. The monthly percentage change of search volumes of each segment during the study period was estimated by the following equation:

$$\log(V_t) = \beta_0 + \beta_1 \times t \ (1)$$

where $\log(V_t)$ is the natural log of the average number of relative search volumes at time t; β_0 indicates the intercept of the model; β_1 is the slope of the model, which was later used to estimate the monthly percentage change of search volumes of each segment [20]; and t is the time variable, indicating the month (from January 2004 to November 2019, indicated as 1 to 191).

Results

Figure 2 shows the policy changes and the search volume trends for the keywords of the 3 working conditions. The red line shows overwork trends, the blue line shows working hours trends, and the gray line shows job stress trends. A higher search volume represents relatively higher public awareness of the topic.

Comparing the relative average search volumes of the 3 working conditions between January 2004 and November 2019, we found the highest volume was for working hours (average volumes=37), followed by overwork (average volumes=13) and job stress (average volumes=11). Most of the observed peaks occurred in 2 periods: the first from the middle of 2010 to early 2011, and the second from the middle of 2015 to early 2018. The first period included peaks in search volumes for both overwork and working hours. The first overwork peak—indicated as the filled circle on the red line—appeared in September 2010, following a death of an engineer in January 2010 that was suspected to be an overwork-related event. Then, a dramatic increase and peak for overwork appeared in March 2011—indicated as the inverted triangle on the red line, following media reports on an increase in suspected overwork-related events and the second revision of Taiwan's Recognition Guidelines for Overwork-Related Diseases in December 2010 (dashed line R2). This phenomenon also urged the government to include more occupations (preschool educator, medical personnel, and hotel and motel bedmaker) among those covered by the Labor Standards Act Articles 30 and 32 in order to regulate their normal working hours and overtime hours. This is indicated as the filled square on the blue line in November 2011.

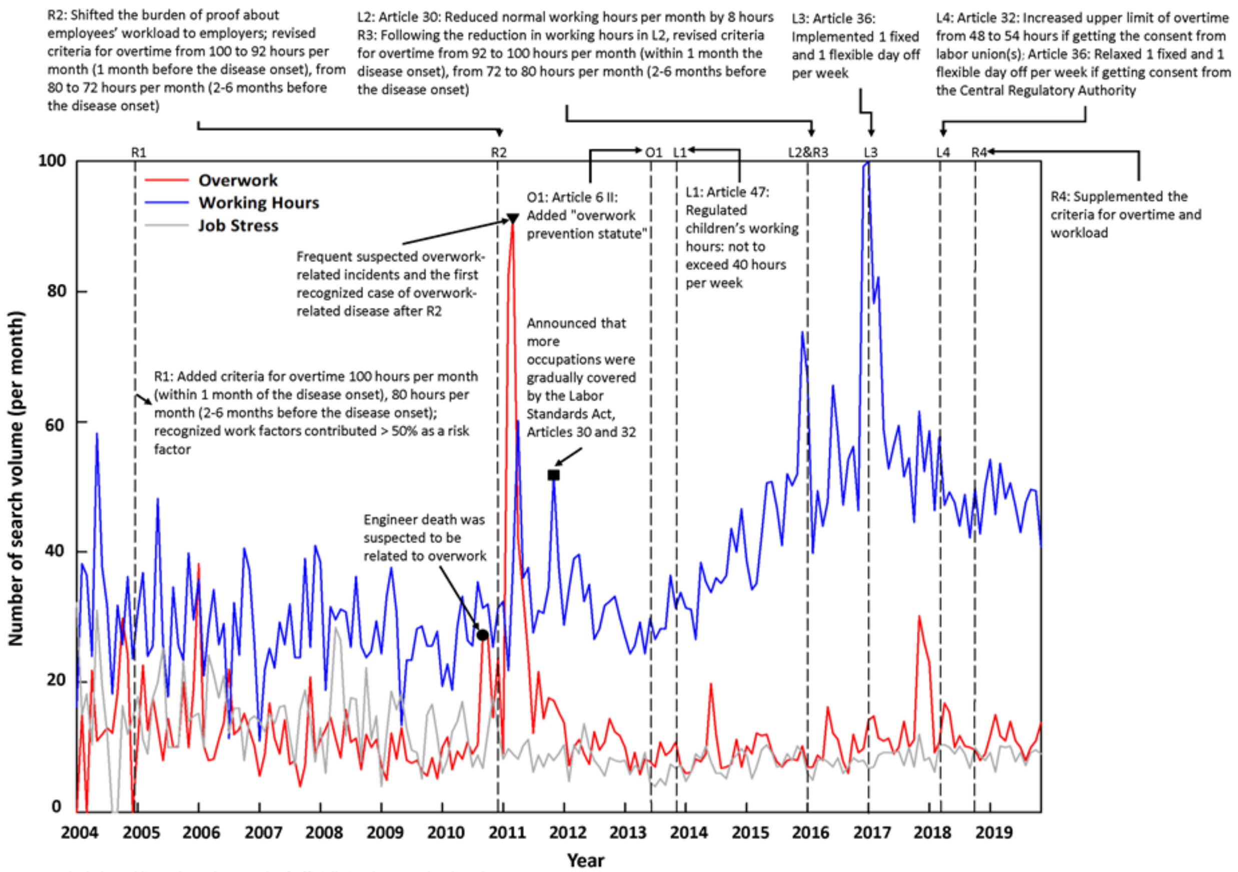
The second peak period appeared to be attributed mainly to policy changes related to working hours: most peaks on working hours searches matched the policy changes, such as the 12th revision of the Labor Standards Act in May 2015, and its implementation in January 2016 (dashed line L2); the 16th amendment of the Labor Standards Act, stipulating 1 fixed and 1 flexible day off per week, in December 2016, and its implementation in January 2017 (dash line L3); and the 18th amendment of the Labor Standards Act, which was implemented in March 2018 (dashed line L4). Over the entire period, the search volume for job stress remained relatively lower and more stable than those for overwork and working hours.

Table 1 presents the monthly percentage change of search volumes for the 3 working conditions, with obvious and significantly increasing trends for both overwork and working hours. From June 2010 to March 2011, the monthly percentage change of search volumes for overwork significantly increased by 28.97% per month (95% CI 19.25-39.49; $P<.001$), following (1) the frequent reports of suspected overwork-related events, especially among engineers and security personnel (Figure 2, inverted triangle), and (2) the second amendment of Recognition Guidelines for Overwork-Related Diseases in 2010 (Figure 2, line R2).

For working hours, the monthly percentage change of search volumes significantly increased from September 2013 to October 2016 by 1.78% per month (95% CI 1.14-2.43; $P<.001$). This trend occurs after the 12th amendment of the Labor Standards Act (ie, reduce 2 hours for normal working hours per week) and the disputes for 16th Labor Standards Act revision (regarding 1 fixed and 1 flexible day off per week); the search volumes peaked in concert with the implementation of the 16th amendment in January 2017.

Slightly increasing trends appeared for job stress from the middle of 2013 to the end of 2019, by 0.39% per month (95% CI 0.05-0.73; $P=.026$), but 84% of the number of search volumes per month was still lower than 10. This trend occurs after the new revision of Occupational Safety and Health Act in July 2013, which added the "overwork prevention statute."

Figure 2. Policy changes and search volume trends for working hours, overwork, and job stress from January 2004 to November 2019.



R2: Shifted the burden of proof about employees' workload to employers; revised criteria for overtime from 100 to 92 hours per month (1 month before the disease onset), from 80 to 72 hours per month (2-6 months before the disease onset)

L2: Article 30: Reduced normal working hours per month by 8 hours
R3: Following the reduction in working hours in L2, revised criteria for overtime from 92 to 100 hours per month (within 1 month the disease onset), from 72 to 80 hours per month (2-6 months before the disease onset)

L3: Article 36: Implemented 1 fixed and 1 flexible day off per week

L4: Article 32: Increased upper limit of overtime from 48 to 54 hours if getting the consent from labor union(s); Article 36: Relaxed 1 fixed and 1 flexible day off per week if getting consent from the Central Regulatory Authority

O1: Article 6 II: Added "overwork prevention statute"

L1: Article 47: Regulated children's working hours: not to exceed 40 hours per week

R4: Supplemented the criteria for overtime and workload

Frequent suspected overwork-related incidents and the first recognized case of overwork-related disease after R2

Announced that more occupations were gradually covered by the Labor Standards Act, Articles 30 and 32

R1: Added criteria for overtime 100 hours per month (within 1 month of the disease onset), 80 hours per month (2-6 months before the disease onset); recognized work factors contributed > 50% as a risk factor

Engineer death was suspected to be related to overwork

Black dotted lines show the month of officially implemented policy changes
(R: revisions of Recognition Guidelines for Overwork-Related Diseases, O: revisions of the Occupational Safety and Health Act, L: revisions of the Labor Standards Act)

Table 1. Trends and monthly percentage change of search volumes for the 3 working conditions.

Working conditions and their periods studied	Trend (*P* value)	Monthly percentage change of search volumes (95% CI)
Overwork		
January 2004 to June 2010	↓ (*P*<.001)	–1.08 (–1.45 to –0.71)
June 2010 to March 2011	↑ (*P*<.001)	28.97 (19.25 to 39.49)
March 2011 to July 2011	↓ (*P*=.014)	–36.06 (–55.14 to –8.87)
July 2011 to November 2015	↓ (*P*=.004)	–0.99 (–1.66 to –0.32)
November 2015 to December 2017	↑ (*P*=.006)	2.99 (0.86 to 5.16)
December 2017 to November 2019	↓ (*P*=.020)	–2.50 (–4.55 to –0.41)
Working hours		
January 2004 to September 2013	↔ (*P*=.768)	0.02 (–0.11 to 0.15)
September 2013 to October 2016	↑ (*P*<.001)	1.78 (1.14 to 2.43)
October 2016 to January 2017	↔ (*P*<.470)	19.07 (–26.00 to 91.59)
January 2017 to May 2017	↔ (*P*=.194)	–14.22 (–32.00 to 8.22)
May 2017 to November 2019	↔ (*P*=.121)	–0.59 (–1.34 to 0.16)
Job stress		
January 2004 to January 2008	↓ (*P*<.001)	–0.93 (–1.45 to –0.41)
January 2008 to April 2008	↔ (*P*=.810)	12.49 (–57.19 to 195.55)
April 2008 to March 2013	↓ (*P*<.001)	–1.64 (–2.10 to –1.18)
March 2013 November 2019	↑ (*P*=.026)	0.39 (0.05 to 0.73)

Discussion

Principal Findings

Over the past 15 years, Taiwanese authorities have revised and implemented several policies to prevent overwork-related health problems. By treating search volumes for overwork-related keywords as proxy indicators of public awareness, our findings suggest that increasing public awareness of the importance of overwork prevention and shorter working hours was observed after Taiwan's recent policy changes related to overwork prevention. Such increased awareness is expected to contribute to better attitudes and practices regarding overwork prevention among the general public [9].

Public awareness of overwork significantly increased ($P<.001$) following the 2010 revision and implementation of the Recognition Guidelines for Overwork-Related Diseases, as well as the frequent media reports in 2009–2011 of suspected overwork-related incidents. Public awareness of working hours increased from 2013 and peaked twice, in the end of 2015 and early 2017, along with the two amendments (12th and 16th) of the Labor Standards Act.

Overwork

The search volumes for overwork increased from June 2010 to March 2011, a period which saw frequent reports of suspected overwork-related incidents prompt public realization of the seriousness of the problem. During the same period, more overwork cases were recognized by the 2010 Recognition Guidelines for Overwork-Related Diseases, which encouraged public confidence in overwork-related policy. After the revised recognition policy was implemented in 2010, the average number of working hours per month of those employed in the industry and service sectors in Taiwan decreased by 9.1 hours from 178.7 hours in 2011 to 169.6 hours in 2016 [21]. The amendment to the recognition guidelines followed the sudden death, in January 2010, of an engineer in Taiwan by cardiogenic shock and hypertrophic cardiomyopathy. Although some of the disease cases could not be officially recognized as related to overwork right out, they exerted influences because a nongovernmental organization (Taiwan Labour Front) and a legislator (Ms Sue‐Ying Huang) organized several press conferences on behalf of families of deceased to promote social awareness via social media, and finally intervene in the decision making of the Ministry of Labor in Taiwan [6]. These activities had aroused public awareness about the strenuous working conditions of elite image engineers. When the 2010 Recognition Guidelines for Overwork-Related Diseases were implemented, cerebrovascular and cardiovascular diseases related to working conditions with long-term excessive overtime became the first recognized overwork disease [6]. Social media together with efforts of Legislator Huang and nongovernmental organizations appeared to have a strong influence on the policy changes in Taiwan [6,22]. Based on this understanding, one way to significantly enhance public awareness of overwork is to cooperate with legislators and nongovernmental organizations and to encourage more media reports on policy changes via social media.

Working Hours

Public awareness about working hours significantly increased ($P<.001$) from September 2013 and reached a peak in January 2017, which included 2 relevant policy amendments to the Labor Standards Act. Among all search volume peaks in this period, the highest were observed in January 2017, aligned with the start of implementation of the 16th amendment, which called for "1 fixed and 1 flexible day off per week." That this policy change would call forth such a high search volume might be due to presidential influence. The theme of overwork was a hot issue during Taiwan's 2016 presidential election. Among all presidential candidates, one candidate (Dr Ing-Wen Tsai, who later won the election) was most committed to preventing overwork problem in Taiwan and even claimed to have prepared related policies that would improve working conditions. This candidate went on to get most of the labor unions' support. Along with their candidate's victory in the presidential election, union workers now look forward to the corresponding policy changes [23], which they expect to provide better working conditions. Since President Tsai was inaugurated in May 2016, the Labor Standards Act has been revised seven more times between November 2016 and May 2019 [24].

Policies and Public Awareness

In our study, we observed higher search volumes around the months when policies were revised and implemented. Revision and implementation are the 2 phases wherein the public can participate in the policy change process (comprised from invention, estimation, selection [revision], implementation, evaluation, to termination) [25]. High search volume observed during policy revision and implementation periods may be attributable to the healthy democracy in Taiwan, which allows for public participation in the policy change process, promotes mutual understanding among stakeholders, and emphasizes the legitimacy of governance [26].

Public awareness on both overwork and working hours was enhanced along with the related policy changes and frequent reports of suspected overwork-related events. However, the search volume trends on overwork and working hours differ. The trend of increasing search volumes on overwork (June 2010 to March 2011) lasted for only 10 months, following the reporting of suspected overwork-related events. By contrast, the trend of increasing search volumes on working hours (September 2013 to October 2016) lasted for 38 months, following the amendments of the Labor Standards Act. Our results suggest that public awareness caused by events is short lived, whereas public awareness induced by policy exists for a longer time, likely because social media continues to provide new information and discussion on the topic [27,28]. Therefore, we suggest that developing overwork-related policies that clearly address risk factors contributes to increasing the public awareness of overwork prevention.

As a reference outcome, public awareness of job stress was lower than public awareness of overwork and working hours over the past 15 years. Given the fact that high job stress is one of the major reasons for overwork [29], we suggest that policies should also target overwork caused by job stress, similar to

European policies for psychosocial risk management in the workplace [30].

Limitations

There are some limitations to our study. First, we identified one keyword with the highest search volume as the representative keyword for each working condition. The general public might use other unpopular keywords (synonyms) to search for more knowledge on each working condition. That is, each working condition could be described using different keywords. Thus, our choice of high-volume keywords might underestimate the effect of policy changes on public awareness.

Second, we could not get an absolute value for search volume on Google Trends. Data obtained from Google Trends are a relative search volume (search volume range from 0 to 100). For cross-keyword and overtime comparisons, we could only compare search volumes for up to 5 keywords at once. Our choice of 3 keywords ensured the comparability of search volumes in our results. Future studies that attempt to expand the number of keywords compared will warrant the development of advanced adjustment methods.

Third, search volumes on Google Trends have been used to explore and predict behaviors on the outpatient visits and people's awareness of health care [16,31-33], but have been shown to be less predictive of public opinions on elections and referendums in low-internet-use or low-freedom-of-speech regions [34]. Our study focuses on searches in Taiwan, where over 80% of the population uses the internet and people have a high degree of freedom to speak their personal opinions [35,36]. The advantages of using Google Trends should also be highlighted, including reducing the time and cost associated with the traditional measures of public awareness and acting as a supplement and complement to traditional polls or surveillance systems [37], especially in countries with high internet penetration and freedom of speech. Although Google is the most commonly used search engine in Taiwan, there are other types of search engines that may be used by different sectors of population, thus may generate different search patterns. Yet, comparative analyses of search patterns across different search engines are beyond the scope of this study. A separate study involving the collection and comparison of data from different search engines can help explore the diverse patterns.

As described in a 2014 review about big data's features [38], the 4V concept—related to the high volume, velocity, variety, and veracity of information—was developed, along with the understanding that transforming data into value requires specific technology and analytical methods. Search volumes obtained from Google Trends are relative, and some might doubt if these search volumes should be considered big data. However, Google Trends data fit the aforesaid 4 features, and some earlier studies have set precedents for considering Google Trends data as online big data [39]. Moreover, the novel way in which Google Trends presents data largely reduces equipment requirements for big data analysis and can be used in many fields. Google Trends can in fact help government officials quickly clarify fast-changing events that raise public concern [40,41] and collect large volumes and a variety of past data in a short period, which is hard to achieve through traditional public polls.

What our study contributes to the global society, particularly labor policymakers, is that a clear policy will lead to rising public awareness, and such an increase in public awareness will reinforce policy change—leading to a virtuous circle. In addition, study findings from Taiwan, a democratic society, highlight the importance of communication capacity in informing and engaging citizens. We believe that our approach enabled a reasonable representation of countries with similar internet access, freedom of participation in political and public affairs, and government responsiveness; by contrast, countries with different political environments warrant separate analyses. We encourage the country's citizens to have more access and freedom to engage in public affairs, particularly labor and public health issues.

Conclusions

Overwork-related health problems due to long working hours, heavy workloads, hectic work pace, and insufficient rest time are major occupational health concerns in many East Asian countries, and in certain occupational groups of Western societies. Regulations concerning working hours, overtime pay, paid leave, and rest time have long been central issues in labor and health protection policies. While in Taiwan, the labor and health authorities have made efforts by revising laws and implementing new policies to improve work quality and to reduce overwork-related health risks, there are still controversies and debates with respect to their regulatory designs and enforcement. Analyses of substantive policies are beyond the scope of this study.

Yet, this study demonstrates rising public awareness of overwork prevention, particularly as it relates to overwork and working hours, following recent overwork-related policy changes. Public awareness of policy contents and citizens' supports are essential for effective policy implementation. Our findings suggested that in Taiwan, certain components of overwork prevention policies were conveyed to the public more effectively while the others were conveyed less effectively. Because our findings indicate that the public are less aware of the role of job stress in overwork prevention, we suggest implementing interventions to address this phenomenon. Ultimately, to enhance public awareness of overwork prevention, national overwork-related policies that clearly address risk factors are necessary.

Acknowledgments

We acknowledge the support from the Ministry of Science and Technology, Taiwan (MOST 107-2314-B-039-062-MY3 and MOST 108 2813-C-039-121-B) and China Medical University (CMU108-MF-09 and CMU108-SR-94). We thank the editors at Tandem Editing LLC and Editage for their careful English language editing.

Authors' Contributions

R-TL contributed to idea formulation, study design, data interpretation, and writing of the manuscript. YC contributed to writing of the manuscript. Y-CJ contributed to the literature review, data collection, data analysis, data interpretation, reporting of results, and participated in the first draft of the manuscript.

Conflicts of Interest

None declared.

References

1. Uehata T. Long working hours and occupational stress-related cardiovascular attacks among middle-aged workers in Japan. J Hum Ergol 1991 Dec;20(2):147-153. [Medline: 1842961]
2. Ke DS. Overwork, stroke, and karoshi-death from overwork. Acta Neurol Taiwan 2012 Jun;21(2):54-59 [FREE Full text] [Medline: 22879113]
3. Lee DW, Hong YC, Min KB, Kim TS, Kim MS, Kang MY. The effect of long working hours on 10-year risk of coronary heart disease and stroke in the Korean population: the Korea National Health and Nutrition Examination Survey (KNHANES), 2007 to 2013. Ann Occup Environ Med 2016 Nov;28:64 [FREE Full text] [doi: 10.1186/s40557-016-0149-5] [Medline: 27891238]
4. Nishiyama K, Johnson JV. Karoshi--death from overwork: occupational health consequences of Japanese production management. Int J Health Serv 1997;27(4):625-641. [doi: 10.2190/1JPC-679V-DYNT-HJ6G] [Medline: 9399110]
5. Cheng Y, Park J, Kim Y, Kawakami N. The recognition of occupational diseases attributed to heavy workloads: experiences in Japan, Korea, and Taiwan. Int Arch Occup Environ Health 2012 Oct;85(7):791-799. [doi: 10.1007/s00420-011-0722-8] [Medline: 22109675]
6. Chang HH, Lin RT. Policy changes for preventing and recognizing overwork-related cardiovascular diseases in Taiwan: An overview. J Occup Health 2019 Jul;61(4):278-287 [FREE Full text] [doi: 10.1002/1348-9585.12046] [Medline: 30816617]
7. Kim I, Koo MJ, Lee HE, Won YL, Song J. Overwork-related disorders and recent improvement of national policy in South Korea. J Occup Health 2019 Jul;61(4):288-296 [FREE Full text] [doi: 10.1002/1348-9585.12060] [Medline: 31025505]
8. Kivimäki M, Jokela M, Nyberg ST, Singh-Manoux A, Fransson EI, Alfredsson L, IPD-Work Consortium. Long working hours and risk of coronary heart disease and stroke: a systematic review and meta-analysis of published and unpublished data for 603,838 individuals. Lancet 2015 Oct 31;386(10005):1739-1746 [FREE Full text] [doi: 10.1016/S0140-6736(15)60295-1] [Medline: 26298822]
9. Badran IG. Knowledge, attitude and practice the three pillars of excellence and wisdom: a place in the medical profession. Eastern Mediterranean Health Journal 1995;1(1):8-16.
10. Scantlebury A, Booth A, Hanley B. Experiences, practices and barriers to accessing health information: A qualitative study. Int J Med Inform 2017 Jul;103:103-108. [doi: 10.1016/j.ijmedinf.2017.04.018] [Medline: 28550995]
11. Health Online. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-01-19]
12. Search Engine Market Share in Taiwan - December 2019. URL: https://gs.statcounter.com/search-engine-market-share/all/taiwan [accessed 2020-01-19]
13. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015 Jun 08;5(6):e006678. [doi: 10.1136/bmjopen-2014-006678]
14. Wang J, Athanasopoulos G, Hyndman RJ, Wang S. Crude oil price forecasting based on internet concern using an extreme learning machine. International Journal of Forecasting 2018 Oct;34(4):665-677. [doi: 10.1016/j.ijforecast.2018.03.009]
15. Mahroum N, Bragazzi NL, Sharif K, Gianfredi V, Nucci D, Rosselli R, et al. Leveraging Google Trends, Twitter, and Wikipedia to Investigate the Impact of a Celebrity's Death From Rheumatoid Arthritis. J Clin Rheumatol 2018 Jun;24(4):188-192. [doi: 10.1097/RHU.0000000000000692] [Medline: 29461342]
16. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]
17. Yamauchi T, Yoshikawa T, Takamoto M, Sasaki T, Matsumoto S, Kayashima K, et al. Overwork-related disorders in Japan: recent trends and development of a national policy to promote preventive measures. Ind Health 2017 Jun 08;55(3):293-302 [FREE Full text] [doi: 10.2486/indhealth.2016-0198] [Medline: 28154338]
18. Google Trends. URL: https://trends.google.com.tw/trends/?geo=TW [accessed 2020-01-19]
19. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014 Oct 22;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
20. Kim H, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med 2000 Feb 15;19(3):335-351. [doi: 10.1002/(sici)1097-0258(20000215)19:3<335::aid-sim336>3.0.co;2-z] [Medline: 10649300]
21. Ministry of Labor. MOL Statistic. URL: https://statfy.mol.gov.tw/statistic_DB.aspx [accessed 2020-02-10]

22. Cheng Y. Policy responses to work-related stress: Examining Taiwan’s experiences from a welfare state regime perspective. Safety Science 2015 Oct;78:111-116. [doi: 10.1016/j.ssci.2015.04.018]
23. Stimson JA. Public Support for American Presidents: A Cyclical Model. Public Opinion Quarterly 1976 Mar 01;40(1):1-21. [doi: 10.1086/268264]
24. Legislative History of Labor Standards Act. URL: https://law.moj.gov.tw/ENG/LawClass/LawHistory.aspx?pcode=N0030001 [accessed 2020-01-19]
25. Brewer GD. The policy sciences emerge: To nurture and structure a discipline. Policy Sci 1974 Sep;5(3):239-244. [doi: 10.1007/bf00144283]
26. Michels A, De Graaf L. Examining Citizen Participation: Local Participatory Policy Making and Democracy. Local Government Studies 2010 Aug;36(4):477-491. [doi: 10.1080/03003930.2010.494101]
27. Anthony D. Up and down with ecology: the issue-attention cycle. The Public Interest 1972;28(1):37-50.
28. Chiu APY, Lin Q, He D. News trends and web search query of HIV/AIDS in Hong Kong. PLoS ONE 2017 Sep 18;12(9):e0185004. [doi: 10.1371/journal.pone.0185004]
29. Viviers S, Lachance L, Maranda MF, Ménard C. Burnout, psychological distress, and overwork: the case of Quebec's ophthalmologists. Can J Ophthalmol 2008 Oct;43(5):535-546. [doi: 10.3129/i08-132] [Medline: 18982028]
30. Leka S, Jain A, Iavicoli S, Vartia M, Ertel M. The role of policy for the management of psychosocial risks at the workplace in the European Union. Safety Science 2011 Apr;49(4):558-564. [doi: 10.1016/j.ssci.2010.02.002]
31. Kaleem T, Malouff T, Stross W, Waddle M, Miller D, Seymour A, et al. Google Search Trends in Oncology and the Impact of Celebrity Cancer Awareness. Cureus 2019 Aug 10;11(8):e5360 [FREE Full text] [doi: 10.7759/cureus.5360] [Medline: 31608195]
32. Hao Z, Liu M, Ge X. Evaluating the impact of health awareness events on Google search frequency. Prev Med Rep 2019 Sep;15:100887 [FREE Full text] [doi: 10.1016/j.pmedr.2019.100887] [Medline: 31194106]
33. Jabaley CS, Blum JM, Groff RF, O'Reilly-Shah VN. Global trends in the awareness of sepsis: insights from search engine data between 2012 and 2017. Crit Care 2018 Jan 17;22(1):7 [FREE Full text] [doi: 10.1186/s13054-017-1914-8] [Medline: 29343292]
34. Mavragani A, Tsagarakis KP. Predicting referendum results in the Big Data Era. J Big Data 2019 Jan 14;6(1):3. [doi: 10.1186/s40537-018-0166-z]
35. Freedom in the World. URL: https://freedomhouse.org/report/freedom-world/2019/taiwan [accessed 2020-01-19]
36. Internet Use in Taiwan. URL: https://report.twnic.tw/2018/TrendAnalysis_internetUsage_en.html [accessed 2020-01-19]
37. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168 [FREE Full text] [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]
38. Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform 2014 May 22;9:8-13 [FREE Full text] [doi: 10.15265/IY-2014-0024] [Medline: 24853034]
39. Ward J, Barker A. Undefined by data: a survey of big data definitions. ArXiv Preprint 2013:1309.5821 [FREE Full text]
40. Ma-Kellams C, Bishop B, Zhang MF, Villagrana B. Using "Big Data" Versus Alternative Measures of Aggregate Data to Predict the U.S. 2016 Presidential Election. Psychol Rep 2018 Aug;121(4):726-735. [doi: 10.1177/0033294117736318] [Medline: 29298630]
41. Fang Z, Chen CC. A novel trend surveillance system using the information from web search engines. Decision Support Systems 2016 Aug;88:85-97. [doi: 10.1016/j.dss.2016.06.001]

Edited by G Eysenbach; submitted 10.02.20; peer-reviewed by D Carvalho, S Van Poucke; comments to author 30.03.20; revised version received 03.04.20; accepted 14.05.20; published 26.06.20.

Please cite as:
Lin RT, Cheng Y, Jiang YC
Exploring Public Awareness of Overwork Prevention With Big Data From Google Trends: Retrospective Analysis
J Med Internet Res 2020;22(6):e18181
URL: https://www.jmir.org/2020/6/e18181
doi:10.2196/18181
PMID:32589160

Original Paper

Impact of Age on Patients' Communication and Technology Preferences in the Era of Meaningful Use: Mixed Methods Study

Martina A Clarke[1], PhD; Ann L Fruhling[2], PhD; Marilyn Sitorius[3], MD; Thomas A Windle[1], BSc; Tamara L Bernard[1], MPH, RN; John R Windle[1], MD

[1]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, United States

[2]School of Interdisciplinary Informatics, College of Information Science and Technology, University of Nebraska Omaha, Omaha, NE, United States

[3]College of Public Health, University of Nebraska Medical Center, Omaha, NE, United States

Corresponding Author:
Martina A Clarke, PhD
Division of Cardiovascular Medicine
Department of Internal Medicine
University of Nebraska Medical Center
982265 Nebraska Medical Center
Omaha, NE, 68198-2265
United States
Phone: 1 402 559 4000
Email: martina.clarke@unmc.edu

Abstract

Background: Identifying effective means of communication between patients and their health care providers has a positive impact on patients' satisfaction, adherence, and health-related outcomes.

Objective: This study aimed to identify the impact of patients' age on their communication and technology preferences when managing their health. We hypothesize that a patient's age affects their communication and technology preferences when interacting with clinicians and managing their health.

Methods: A mixed methods study was conducted to identify the preferences of patients with cardiovascular diseases. Results were analyzed based on the patients' age. Grounded theory was used to analyze the qualitative data. Patients were recruited based on age, gender, ethnicity, and zip code.

Results: A total of 104 patients were recruited: 34 young adults (19-39 years), 33 middle aged (40-64), and 37 senior citizens (>65). Young adults (mean 8.29, SD 1.66) reported higher computer self-efficacy than middle-aged participants (mean 5.56, SD 3.43; $P<.05$) and senior citizens (mean 47.55, SD 31.23; $P<.05$). Qualitative analysis identified the following three themes: (1) patient engagement (young adults favored mobile technologies and text messaging, middle-aged patients preferred phone calls, and senior citizens preferred direct interactions with the health care provider); (2) patient safety (young adults preferred electronic after-visit summaries [AVS] and medication reconciliation over the internet; middle-aged patients preferred paper-based or emailed AVS and medication reconciliation in person; senior citizens preferred paper-based summaries and in-person medication reconciliation); (3) technology (young adults preferred smartphones and middle-aged patients and senior citizens preferred tablets or PCs). Middle-aged patients were more concerned about computer security than any other group. A unique finding among senior citizens was the desire for caregivers to have access to their personal health record (PHR).

Conclusions: Patients of different ages have different communication and technology preferences and different preferences with respect to how they would like information presented to them and how they wish to interact with their provider. The PHR is one approach to improving patient engagement, but nontechnological options need to be sustained to support all patients.

(J Med Internet Res 2020;22(6):e13470) doi:10.2196/13470

KEYWORDS

patients; personal health record; needs assessment; data display; communication; age groups

Introduction

Background

The purpose of the Health Information Technology for Economic and Clinical Health Act was to promote the adoption and meaningful use of health information technology (HIT) [1]. Meaningful use is an endeavor initiated by the Centers for Medicare and Medicaid Services (CMS) and the Office of the National Coordinator for Health IT. It involves the use of certified electronic health records (EHR) fundamentally to improve the quality, safety, and efficiency of care [2]. One of the meaningful use requirements is to engage patients and families in their health by offering secure, Web-based access to patients' health information and provide tools that support electronic communication between patients and providers [3]. A more robust understanding of patient communication and technology preferences is essential for ensuring that patients have access to relevant information needed to make informed health care decisions [4-7]. Providing patients access to personal health information has the potential to enhance the delivery of health care services and encourage patient engagement [8]. Patient experience is an integral component of health care quality and focuses on aspects of health care delivery, such as getting timely appointments, easy information access, and effective communication with health care providers. Accurate perception of the patient experience is an essential step in moving toward patient-centered care [4,9]. Patient-centered care is 1 of the 6 major domains of health care quality and is defined as "providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patients' values guide all clinical decisions [10]." Identifying effective means of communication between patients and their health care providers have a positive impact on patients' satisfaction, adherence, and health-related outcomes [11-13]. This allows for critical and necessary communication to enable the highest level of patient-centered care possible.

Electronic health (eHealth) is an effective medium for bringing HIT to patients in ways that are easily accessible [14]. eHealth is "*health services and informational resources delivered or enhanced through the Internet and related information and communication technologies* [15]." eHealth supports communication between health care providers and their patients and promotes patients' self-management of chronic diseases [16-19]. As the use of eHealth increases, it is crucial to examine sociodemographic characteristics, such as age, that may affect patients' ability to use eHealth [20-26]. Younger adults are historically more open to using eHealth than older adults [25,27-35]. Not all older adults own internet-connected devices to access eHealth tools or have the necessary skills needed to use it, increasing the digital divide gap [27,32,33,35].

With increased access to the internet, digital divide research has shifted from internet connectivity barriers (level one of the digital divide) to barriers related to internet skills and usage (level two of the digital divide) [36]. The third level of the digital divide focuses on individuals' capacity to translate their internet access and use into favorable tangible outcomes [37-39]. The digital divide has the potential to intensify health disparities among vulnerable populations; previous research provides evidence of the digital divide being a barrier to eHealth adoption [24,26,40-49]. Providing effective eHealth interventions is an opportunity to provide patients with the necessary communication tools for self-management [25,50,51]. This opportunity could be easily missed if patients' needs, preferences, and abilities are neglected [14].

Objectives

This study's objective was to identify the impact of patients' age on their communication and technology preferences when managing their health. Communication and trust are at the core of the patient-clinician interaction. In addition, technology adoption outside of health care is heavily influenced by a user's age [52]. Thus, we investigated the communication and technology preferences of patients with cardiovascular diseases in the age of CMS' Promoting Interoperability Program. Dholakia [52] presents a general framework that an individual's age is a critical determinant of a person's attitude, use, and perceived benefits of many technologies. Age is influential in technology use and adoption because the periods of a person's life, notably during years as a youth, shape the skills, orientations, and environments of particular technologies [52]. Therefore, we hypothesized that communication and technology preferences are determined, in part, by the age of the patient. The study aimed to test this hypothesis with the following research question: How does age impact patients' preferences when managing their health?

Methods

Study Design

A mixed methods approach was applied to examine communication and technology preferences by conducting interviews with 104 cardiovascular medicine patients at the University of Nebraska Medical Center (UNMC) in the United States of America. UNMC's Institutional Review Board approved the study.

Organizational Setting

UNMC is a Midwest academic medical center whose clinical partner is Nebraska Medicine. The Division of Cardiovascular Medicine operates three clinics with over 28,000 annual patient visits. The Health care Information and Management Systems Society, a nonprofit organization that examines hospitals' advancement toward implementing HIT, has honored UNMC with Stage 7 of the Electronic Medical Record Adoption Model [53]. Stage 7 means UNMC's hospitals and clinics have a fully integrated EHR, employ Continuity of Care Documents to transport data, use data warehousing to assess clinical data, and demonstrate summary data continuity for all hospital services [54]. The personal health record (PHR) offered at UNMC's is Epic MyChart, a tethered PHR.

Participants

Patients with heart disease were chosen for this study because cardiovascular disease is a complex, chronic disease and is one of the leading causes of death in the United States. Heart disease and stroke account for over 800,000 deaths annually, costing

US $316 billion in health care expenditures and lost productivity yearly [55]. Although patients were recruited from the cardiovascular clinic, patients in this sample had comorbidities. A cardiovascular medicine research nurse coordinator recruited patients after their clinic visit based on age, gender, race, ethnicity, and zip code. Zip codes were used to determine urban versus rural environments and establish average household income. Rural areas were defined as a population less than 10,000 and nonadjacent to a metropolitan area. Income was based on census data of median household income for that zip code. Participants were grouped according to their age. Younger participants were categorized as *young adults* (19-39), senior citizens in the US begin at age 65 and were categorized as *senior citizens* in this study, and middle aged were categorized as adults in between the age ranges of the young adults and senior citizens (40-64). Data collection sessions were scheduled and conducted in a clinic or adjacent conference room. Whenever possible, the research interview was linked to their scheduled appointment. Participants were not compensated for their participation.

Data Collection

Each data collection session lasted 30 to 60 min. After the consent was obtained, sociodemographic data were obtained, and instruments to measure health literacy, confidence using technology, and patient activation were administered followed by the interview. Health literacy was measured using three health literacy screening questions for detecting patients with inadequate or marginal health literacy using a 5-point Likert scale [56]. Confidence using technology was measured using the Computer Self-Efficacy Survey, which is a 10-item survey using a 10-point Likert scale to measure a person’s ability to use a computer [57]. Patient activation was measured using the Patient Activation Measure (Insignia Health), which is a 13-question survey using a 4-point Likert scale to measure the level of patients’ engagement in their health [58].

Participants were interviewed using a semistructured interview guide as a part of the grounded theory approach. The session intended to have an open dialog with patients to understand how they currently receive their clinical care. The moderator discussed with the participants a typical clinic visit asking exploratory questions about their care, communication preferences, use of technology, and patient portal, use if any. Members of the research team asked relevant follow-up questions based on participants’ responses. Sampling continued until saturation was reached and no new information was gained during interviews. Study timeframe was from July 2015 to November 2018.

Data Analysis

A mixed methods approach was used to collect and analyze data. Survey data were recorded in a secure database and analyzed in conjunction with a biostatistician using SAS 9.4. Descriptive statistics, which included counts and percentages, means, standard deviations, medians, minimums, and maximums were used to summarize the data. Patient demographics were compared between age categories using the Fisher exact test for categorical data. Comparisons of scores between age groups were done using the Kruskal-Wallis test. Pairwise comparisons were adjusted using Bonferroni method if the overall P value from the Kruskal-Wallis test was significant. All tests were two-sided and a $P<.05$ was considered statistically significant.

The audio-recorded interviews were stored on a secure server. Grounded theory approach was used to analyze the qualitative data. Concurrent data collection and analysis were conducted through various stages of coding. Audio files were transcribed, and two independent, qualitative coders analyzed the transcribed interviews. A codebook was created, and themes were revised until coders came to a consensus. A third independent reviewer settled any disagreements between the two coders. Qualitative data were analyzed using NVivo 11, a qualitative data analysis software (QSR International, Doncaster, Australia).

Results

Participant Characteristics

Table 1 shows the demographics of patients with cardiovascular diseases who participated in the study and whose ages ranged from 19 to 95 years. Responses are classified according to age groups: young adults: 19-39 years of age (34 participants), middle aged: ages 40-64 (33 participants), and senior citizens’ age >65 (37 participants). A difference was found among race across the three age groups.

There is no significant difference among the three age groups’ health literacy ($P=.10$) and activation level ($P=.09$). There is a significant difference between at least 2 of the 3 age groups in computer self-efficacy (Table 2). Young adults report higher computer self-efficacy than middle-aged participants ($P<.05$) and senior citizens ($P<.05$).

Table 1. Demographics of 104 patients with cardiovascular diseases who participated in interviews presented as percentages. Examined demographics include gender, race, age, location, income, and personal health record use.

Demographics	Aged 19-39 years, n (%)	Aged 40-64 years, n (%)	Aged >65 years, n (%)	*P* value
Age (mean 55 years)	34 (33)	33 (32)	37 (36)	
Gender				.49
Male	14 (41)	19 (56)	18 (53)	
Female	20 (59)	14 (41)	19 (56)	
Race and ethnicity				.02
African American	3 (9)	10 (29)	4 (12)	
Hispanic	6 (18)	1 (3)	1 (3)	
White	24 (71)	22 (65)	30 (88)	
Asian	1 (3)	0 (0)	0 (0)	
American Indian	0 (0)	0 (0)	2 (6)	
Location by zip codes				.14
Rural (population<10,000)	11 (32)	7 (21)	12 (35)	
Urban (population>10,000)	23 (68)	26 (76)	25 (74)	
Income by zip codes				.69
Average household income <US $45,000	5 (15)	12 (35)	10 (29)	
Average household income >US $45,000	29 (85)	21 (62)	27 (79)	
Personal health record user				.55
Yes	26 (76)	17 (50)	19 (56)	
No	8 (24)	16 (47)	18 (53)	

Table 2. Comparison of composite scores of surveys between age groups.

Survey	Count, n	Mean (SD)	Median (min-max)	*P* value[a]
Computer self-efficacy				<.001
Young adult	34	8.29 (1.66)	9 (3.6-10)	
Middle aged	30	5.56 (3.43)	5.85 (0-10)	
Senior	31	4.78 (3.13)	4.6 (0-9.9)	
Patient activation measure				.09
Young adult	34	72.18 (14.43)	71.34 (51-100)	
Middle aged	30	63.4 (15.7)	59.38 (44-100)	
Senior	31	65.24 (20.13)	65.47 (44-100)	
Health literacy				.16
Young adult	34	2.6 (0.42)	2.33 (2-3.7)	
Middle aged	30	2.84 (0.5)	2.67 (2-4)	
Senior	32	2.82 (0.74)	2.67 (1.3-5)	

[a] Pairwise comparisons: young vs middle aged, *P*<.05; young vs senior, *P*<.05.

Impact of Age on Patients' Preferences When Managing Their Health

In total, three themes emerged by analyzing and combining findings to form overarching concepts: Patient Engagement, Patient Safety, and Technology.

Patient Engagement

Patient Engagement is based on the foundations of trust and communication, which are two common and important sub-themes identified by patients across all demographics. Trust and communication in patient-provider relationships can influence patients' communication preferences [59].

Similarities Among Age Groups

Patients view care as continuous rather than episodic (ie, a clinic visit) and expect communication to occur both before and after appointments. Calendar reminders are preferred (paper and electronic) with the option to add appointments directly to their calendar.

> *When she comes to give me my paper, I will go and start putting it in my calendar. I will put it in my calendar on my phone which also pops up in my tablet because it is through google. I will put it on my calendar at work…* [Participant 394-082, age 37]

Between visits, patients who use a PHR prefer to communicate electronically with their physicians via their PHR or email. PHR nonusers prefer to communicate in person or over the phone. Irrespective of age, trust is a common subtheme and is vital to all age groups with one exception; middle-aged patients are less likely to trust technology than young adults or senior citizens. Irrespective of age, most patients prefer to complete questionnaires before their clinic visit either over the phone or electronically.

> *I like it when they send them to you before. I have had that done here, and that is nice. It saves me some time because I don't live in Omaha. I drive like 45 minutes in. So that is nice.* [Participant 394-087, age 35]

Patients have strong negative feelings about redundant questioning during their clinic visits. All patients understand the importance of knowing their diagnosis, lab results, and medications and prefer their physician's impressions, in lay terms, rather than a table of results.

> *Instead of the big words and I have no idea what it is. That would be great.* [Participant 394-030, age 67]

Differences Among Age Groups

Differences in preferences are found between groups (Table 3). Young adults have an interest in communicating with other patients with similar conditions via social groups but strongly prefer those sessions to be moderated by professionals. Young adults also request the ability to update information through the PHR and to communicate with providers through text, email, and secure messaging. Of note, social media is not a priority for middle-aged adults. They prefer to contact their physician by phone or email rather than texting. Senior citizens prefer communication with their providers through phone calls, email, and in person and are not interested in participating through social media.

Table 3. Preferences between age groups (young adults, middle-aged adults, and senior citizens) regarding the theme of patient engagement.

Patient engagement[a]	Differences in preferences between age groups
Phone reminder issues	• Young adults: Ineffective because they rarely check their voicemail. • *Sometimes I get a voicemail, and sometimes I don't check my voicemail until a week after. I've never missed an appointment, but just that I read my texts all the time, so that would be nice.* [Participant 394-017, age 24] • Middle-aged adults: Did not mention issues with phone reminders. • Messages begin too abruptly, so they miss the beginning. • *We older people can't just run and jump to the phone. By the time I pick up the phone, the message is done, and I heard it say 3 o'clock. I heard the 13th. I think that the message that they leave on the phone should wait until somebody picks it up.* [Participant 394-036, age 64] • Senior citizens: Miss reminders because of low ringtones. • *All I have to do is answer. My problem is that a lot of times I don't hear because the ring isn't so loud, and by the time I get to it, the caller has already hung up.* [Participant 394-018, age 84]
Provider communication	• Young adults: Text, secure messaging, email • Middle-aged adults: Phone, email • Senior citizens: In-person, phone, email
Scheduling appointments	• Young adults: In-person after a visit, phone, PHR[b] • Middle-aged adults: Phone, open to using PHR • Senior citizens: Phone
Appointment reminders	• Young adults: Electronic calendar, text messages • Middle-aged adults: Electronic and paper calendar, email • Senior citizens: Phone calls, handwritten notes
Repetition	• Young adults: Dislike repetitive questions during current check-in • Middle-aged adults: Do not mind repetitive questions during check-in • Senior citizens: Approve of repetitive questions during the current check-in process because they believe it confirms that the right patient is being treated. • *I don't mind. Sometimes you will see five or six different people, and they will all ask the same thing. I think that's good. That way if they have to give you something, they are not giving it to the wrong person.* [Participant 394-002, age 80]
Caregivers	• Young adults: Did not mention caregiver access. • Middle-aged adults: Some do not want to share information with children because they want to be in control of their health care. • Senior citizens: Wants caregivers to have Web access to health information. • *M: I have a list of medications I carry with me all the time. She comes to the appointments, and she is my secretary.* • *F[spouse]: That way, I know what is going on. I ask, are there any surprises?* [Participant 394-011, age 73]
Lab results	• Young adults: Did not mention a preference • Middle-aged adults: Phone call, letter • Senior citizens: Letters, email
Social engagement	• Young adults: Had an interest in social groups with patients with similar diseases. • Middle-aged adults and senior citizens: Did not mention an interest in social engagement.

[a] The three age groups categorize the findings. Text in *italics* represents participants' quotes.

[b] PHR: personal health record.

Patient Safety

The quality of health care can be improved by focusing on improving patient safety and is an excellent standard of measure for high-quality health care. Patient safety is the effective prevention and amelioration of the risk of medical-error related adverse events [60].

Similarities Among Age Groups

Medication reconciliation is critical in ensuring proper medication adherence and avoiding patient harm from inaccurate information [61,62]. All patient groups are frustrated that their home medications often are not reflected in the EHR despite attempts at reconciliation.

> *Mine [home medication list] is correct, but for some reason, whenever I go in, they have some that are like from 12 years ago. I don't understand why the doctor's list is so wrong compared to mine.* [Participant 394-017, age 24]

The after-visit summary (AVS) reflects a summary of activities that occurred during the clinic visit including lab results, medications and changes, and treatment plans, as well as patient education. All age groups focus on and prefer changes in treatment plans and medications to be highlighted in their AVS.

Age impacts internet use for health information. Patients commonly identify the Google search engine as their health information source. Patients enter their search and select a link from the populated list. They can occasionally name the websites they access for health information. They also request links to credible sources about their diseases.

> *No, I just kinda Google it and go through anything I can get into. I do not belong to any medical thing. Doctors have told us that is probably not a good idea to do all that but I always do.* [Participant 394-011, age 73]

Age also impacts travel with health information. Participants carry some part of their health information with them while traveling. They are also open to alternative ways of travelling with their health information when out of town in case of an emergency, such as, on their phones or on a card the size of a debit card.

> *Not like I should. I mean, I have the thing on my iPhone that is the emergency alert. I put my conditions on that. I guess I kind of, maybe, but it does not have like a list of your medicines or things like that.* [Participant 394-084, age 35]
>
> *…I got a cut deep in my ear from something, and it bled like crazy forever, all of the ways home. It happened in Hawaii, and all of the ways home, we were changing cotton soaked and I came in, and they put me with the ENT out at Oakview [UNMC clinic]. When I went to the hospital, of course, they wanted to know about me before they do anything and so I had to fill out a bunch of paperwork. It would be nice to be able just to hand them the flash drive. They could then plug it in their computer and sync it up.* [Participant 394-007, age 77]

Differences Among Age Groups

There are differences between preferences for the AVS and medication reconciliation at clinic visits (Table 4). Young adults expressed little interest in the paper version of the AVS, preferring an electronic version. Middle-aged and senior citizens prefer a paper-based AVS.

Young adults and senior citizens are most likely to look up health information on the internet than middle-aged adults. Young adults usually search for information about their condition and the meaning of their lab results. Middle-aged adults search for information about their condition and information on nutrition and weight loss. Senior citizens search for information concerning their medications and the meaning of their lab results.

Young and middle-aged adults report visiting trusted, well-known websites for their health information, citing WebMD and hospital websites, such as Mayo Clinic. Young adults verify the information they receive on the internet by comparing the information provided by multiple sites. Senior citizens determine if the information gathered from the internet is accurate by verifying the information with their physician or pharmacist.

Regarding travelling with health information, some young adults do not want to travel with their health information but are open to doing so in the future when their chronic disease is less controlled. Middle-aged adults and senior citizens travel with their current medication list. Senior citizens also carry other pertinent health information, such as medical device information. Young adults and senior citizens are open to traveling with their health information on their phones. Young and middle-aged adults are less enthusiastic about the option of a health information card because the card is another item they would have to keep track of.

Table 4. Preferences between age groups (young adults, middle-aged adults, and senior citizens) regarding the theme of patient safety.

Patient safety	Differences in preferences between age groups
AVS[a]	- Young adults: Did not find AVS very useful - Preferred an electronic copy - Middle-aged adults: Focused on highlighted sections in AVS. - Preferred paper copies for record-keeping. - Senior citizens: Focused on highlighted sections in AVS. - Preferred paper copies for record-keeping.
Medication reconciliation[b]	- Young adults: Did not mention issues with the medication reconciliation process - Middle-aged adults: Carry medication list to clinic visits. - Senior citizens: Prefer to complete before clinic visit electronically or over the phone - Carry medication list to clinic visits.
Navigation	- Young adults: Did not mention navigational functionality within the hospital. - Middle-aged adults: Did not mention navigational functionality within the hospital. - Senior citizens: Desire navigational functionality within the hospital, reporting getting lost more than once in hospitals.
Traveling with health information	- Young adults: Rarely travel with health information but open to doing so on the smartphones - *When I used to go out in Colorado Springs to visit my boyfriend's family, I would tell his mom everything and give her paperwork. So, that way then, in case something happens to me for all I know, I'll get run over by a bull, you know, eaten by a mountain lion.* [Participant 394-089, age 36] - Middle-aged adults: Travel with their current medication list - Senior citizens: Travel with their current medication list and medical device information
Health information source	- Young adults: Trusts well-known health websites (eg, WebMD) and hospital websites (eg, Mayo Clinic) for their health information - Middle-aged adults: Trusts well-known health websites (eg, WebMD) and hospital websites (eg, Mayo Clinic) for their health information - Senior citizens: Identifies mainly Google as their health information source - *No, I am just on Google and I just type in the medicine's name and look it up and see what the side effects are and stuff like that. They show pictures of what they used to look like, the different brands, and get information on them.* [Participant 394-046, age 72]
Information search topics	- Young adults: Search for information about their condition and the meaning of their lab results - *Vagal tone, yes! So like that word, just, I have got to remember all those things, so it's nice to be able to go on there and you know, read things like that.* [Participant 394-034, age 28] - Middle-aged adults: Search for information about their condition and information on nutrition and weight loss. - Senior citizens: Search for information concerning their medications and the meaning of their lab results.

[a] AVS: after-visit summaries.

[b] The 3 age groups categorize the findings. Text in *italics* represents participants' quotes.

Technology

The technology theme focuses on attitudes, device preferences, and password management.

Similarities Among Age Groups

All age groups are open to using a swipe card similar in size to a debit card for checking into their clinic visit. Participants in all age groups find the tables and graphs in the PHR hard to understand. They also prefer a phone call when receiving abnormal test results to allow a dialog concerning the results.

> *…when sometimes things could be abnormal and need to be discussed, the phone call would be better.* [Participant 394-003, age 65]

They also desire medication information in the PHR.

> *What I don't find useful, when he takes his meds, it usually just has your dosage you take and what it is, the name of it, but it doesn't tell you, and I think it should tell you, because he takes so many meds, why, what is this for. Like today, this is for your cholesterol. That is why you are taking it. Because for him, when he goes from doctor to doctor, he has so many, he forgets, and I'll even forget because he just takes so much, so I do think it should say something like that.* [[Participant 394-037, age 57]

Patients find passwords cumbersome to recall and are open to alternative authentication methods for accessing their PHR. Participants attempt to recall their passwords from memory, but to assist with recall, they keep a notebook with a list of their passwords, or they use the same passwords for all the sites they access. They also confess to sharing their passwords with others.

> *I have basically one password for everything I do, and then every six months, I change it…* [Participant 394-001, age 57]

> *I have a book of passwords at home that if she [spouse] needed to get to it, she could get to it.* [Participant 394-004, age 67]

All age groups desire the ability to have electronic access to laboratory and test results with normal values and their trends in the PHR.

> *The synopsis from the doctor looking at it would be my guess. The option of looking at it just for curiosity, but that wouldn't make as much difference as what the doctor would have to say because unless I'm trained in the field, I wouldn't know what I was looking at anyway.* [Participant 394-013, age 65]

Differences Among Age Groups

Young adults prefer to access their PHR through an app on their smartphones (Table 5). Young adults are willing to complete the check-in process for their clinic visits using their phones and are open to Web-based audiovisual clinic visits. Middle-aged participants prefer a PC or tablet to smartphones. Senior citizens are similar; they prefer a PC or tablet because of the need for a larger font. Middle-aged participants express difficulties navigating the PHR. A unique feature request by senior citizens is the ability to delegate PHR access to caregivers.

Young adults maintain their passwords from memory for fear of misplacing a written list. To mitigate this fear, young adults write their passwords in an encrypted format or keep a list in a notes app on their phone or a computer file. Senior citizens' main frustration with passwords is that they cannot remember them because of the differing password criteria that websites require. Some senior citizens keep their password list in a computer file. Misplacing paper versions of their password list is rarely a concern and even have passwords written on Post-It notes taped to their desks at home.

Table 5. Preferences between age groups (young adults, middle-aged adults, and senior citizens) regarding the theme technology.

Technology[a]	Differences in preferences between age groups
Electronic check-in	• Young adult: Very receptive to complete the check-in process for their clinic visit using their phone. • Middle-aged adults: Not open to complete the check-in process for their clinic visit using their phone. • Senior citizens: Preferred human interaction but were open to complete the check-in process for their clinic visit using their phone. • *I would not do FaceTime. I prefer human to human contact.* [Participant 394-015, age 86] • *I do not think it would be necessary. It might add to it for some people. I think there are some people that work better in that kind of environment, but that is not necessary for me.* [Participant 394-007, age 77]
Web-based clinic visit	• Young adult: Very receptive to the convenience of Web-based audiovisual clinic visits. • *If I were master of the universe, I would have FaceTime appointments. I would have it where I can get the lab done, you know, at my local lab link at the local hospital.* [Participant 394-020, age 35] • Middle-aged adults: Split when it came to Web-based and audiovisual clinic visit option. • Senior citizens: Not receptive to the convenience of Web-based audiovisual clinic visits.
Preference: PHR[b] access device	• Young adult: Smartphone • Middle-aged adults: Tablet, computer • Senior citizens: Tablet, computer
Preference: internet access device	• Young adult: Smartphone • Middle-aged adults: Tablet, computer • Senior citizens: Tablet, computer
Accessibility	• Young adult: No issues • Middle-aged adults: Larger font • Senior citizens: Larger font
Password management	• Young adult: Use the same password, alternate between a few passwords, keep a computer file, Post-it • Middle-aged adults: Notebook, use, the same password, spouse • Senior citizens: Use the same password, notebook, Post-it, saved in the browser, computer file.
Message received receipt	• Young adult: No issues • Middle-aged adults: Do not electronically communicate because they don't know if their provider receives messages • Senior citizens: Confirmation receipt to know that their message is received

[a] The three age groups categorize the findings. Text in *italics* represents participants' quotes.

[b] PHR: personal health record.

Discussion

Summary of Findings

This study demonstrates the importance of the role that age plays in determining communication preferences and technology use, with age having a significant impact on patient preferences in clinical care. There was a difference between race across the three age groups. This difference was because our sample was representative of the population studied. The US Census Bureau estimates Nebraska's population to be 78.6% (1,520,445/1,934,408) White, 11.2% (216,654/1,934,408) Hispanic, 5.1% (98,655/1,934,408) Black or African American, 2.6% (52,229/1,934,408) Asian, and 1.4% (29,016/1,934,408) American Indian.

Principal Findings

Young adults had the highest self-efficacy score among the three age groups. Young adults are most receptive to completing the check-in process for their clinic visit using their phone and would opt for the convenience of Web-based audiovisual clinic visits, prefer smartphone access, and have a high use of the PHRs for convenience. Middle-aged adults had lower computer self-efficacy scores than young adults. Middle-aged adults are blenders of old and newer technologies. They prefer phone and email for provider communication and both electronic as well as paper calendars for appointment reminders. Similar to middle-aged adults, senior citizens had lower computer self-efficacy scores than young adults. Senior citizens in this study prefer phone communication and mention caregivers gaining access to their health information.

Comparison With Prior Work

Young adults grew up in the age when mobile devices exploded onto the market and are very familiar with these devices. Results from a 2015 survey conducted by Pew Research Center, determining technology device ownership, found that 86% of young adults (aged 18-29 years), 83% percent of adults (aged 30-49 years), 58% percent of middle-aged adults, and 30% of senior citizens owned a smartphone [63]. Similarly, a study by Zallman et al [64] found that being 40 years old and under was associated with a 50% to 90% increase in the odds of preferring text messaging over other modes of communication. Based on Dholakia’s framework [52], this explains young adults’ preference for mobile PHR access, text message, and electronic access to appointment reminders and their AVS. With a higher reported computer self-efficacy, future research should analyze the effects of a user-friendly mobile PHR interface on increasing adoption among young adults.

Middle-aged and senior patients’ tablet use can be explained by the simplified, less intimidating interface of tablets. A study by Jayroe and Wolfram [65] found that older adults had an overall positive experience when using a tablet to complete search tasks. Previous literature is rich with information on young adults and senior citizens’ as it relates to technology adoption. However, there is a lack of information on how to best present information and increase PHR adoption for middle-aged adults. Middle-aged adults and senior citizens share multiple similarities. Middle-aged adults and senior citizens may be content with their current modes of communication technology use and may be reluctant to attempt to keep up with the fast-changing technology environment.

Although senior citizens are usually behind when it concerns technology adoption, Pew’s report on Americans’ Internet Access found that senior citizens had the highest rate of change from 2000 to 2015, with more than half (58%) of all adults aged 65 years and older using the internet [66]. In addition, in a study by Gordon and Thornbrook [51], over 75% of senior citizens reported having access to a desktop or laptop and 25% owned a tablet. These results suggest a narrowing of the digital divide; therefore, it is crucial to support the senior citizens, who are open to eHealth, to adopt digital tools that enable better health care self-management. Senior citizens are also more likely to use technology that has been around longer and that are most familiar. A study by Olsen et al [67], surveying age-related differences in overall usage of technologies, found that older adults were more likely to use technologies, such as telephones, answering machines, credit cards, etc [67]. Senior citizens in this study also prefer phone communication for the sake of human interaction. Olsen et al [67] similarly discovered that older adults frequently made phone calls when the phone was an option. These results suggest that a combination of new and old technology is still imperative to satisfy the communication needs and preferences of all patients. Senior citizens are also more likely to be more concerned about their health and, maybe, feeling vulnerable because of their dependence on informal, family caregivers for their health management. This may be one motivation to providing caregivers with access to their patients’ health information [68-70]. Allowing caregivers improved access to manage senior patient’s health information is a useful feature to build into PHR redesign.

Patients search for health information on the internet using search engines without considering the links they are selecting from the search list. This is a patient safety issue because patients may be consuming misleading information. Providing patients with trusted links to websites and educating patients on how to determine if an information source is credible are two ways to reduce the potential of patient harm from misinformation.

Limitations

While this study recruited patients based on age, gender, race, ethnicity, and zip code, a continuing study is necessary to confirm, extend, and refine these results, as there may be other complex reasons that may affect results. The sample consists of patients from one health care system, using one patient portal, with similar chronic diseases. Our results need to be confirmed by a larger multisite study. However, patients receive care at institutions outside of Nebraska Medicine and have experience using PHRs from other manufactures. The purpose of this study is not to predict trends for the future, but to determine patients’ current practice, which can inform providers on the information patients seek and can improve the communication providers have with their patients during clinic visits. In addition, these results can inform vendors of patients’ desired functionality for the design and implementation of PHRs to improve the usability of PHRs.

Conclusions

This study demonstrates that as implemented, mandatory PHR use could increase the digital divide among vulnerable populations. Many patients, especially the elderly and those with low reported computer self-efficacy, find technology a barrier to use. PHR vendors should consider that young adults have a high affinity for electronics. Middle-aged adults are mixers of old and new technologies, and in our sample, many do not trust technology, preferring to interact with humans than with technology. Senior citizens also prefer human interaction but are willing to use technology that is familiar to them. The PHR does not suitably support the communication needs of elderly patients and their caregivers. Patient-centered care needs to support individual patient preferences, which includes nontechnology options.

Clinical Relevance Statement

To achieve substantial, meaningful use, HIT must adapt to the user, rather than forcing the user to adapt to the technology. Patients of different ages have different communication and technology preferences, information display differences, and preferences regarding how they wish to interact with their provider. The PHR does not correctly support caregiver access to health information. It is one approach to improving patient engagement, but to support all patients, sustaining nontechnological options are necessary.

Acknowledgments

This study was funded by an AHRQ grant: Optimizing the Electronic Health Record for Cardiology Care (R-01 HS022110-01A1).

Conflicts of Interest

None declared.

References

1. US Department of Health & Human Services. Washington, DC, USA; 2017 Jun 16. HITECH Act Enforcement Interim Final Rule URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html [accessed 2019-07-20]
2. Centers for Disease Control and Prevention. 2017 Jan 18. Public Health and Promoting Interoperability Programs (formerly, known as Electronic Health Records Meaningful Use) URL: https://www.cdc.gov/ehrmeaningfuluse/introduction.html [accessed 2019-07-20]
3. Centers for Medicare & Medicaid Services. Washington, DC, USA; 2016. Medicaid EHR Incentive Program Stage 3 Patient Electronic Access to Health Information 2016 URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/MedicaidStage3_PatientElectronicAccessTipsheet.pdf [accessed 2017-12-11]
4. Rutten LJ, Agunwamba AA, Greene SM, Mazor KM, Ebbert JO, St Sauver JL, et al. Enabling patient-centered communication and care through health information technology. J Commun Healthc 2014;7(4):255-261. [doi: 10.1179/1753807614y.0000000067]
5. Kish-Doto J, Scales M, Eguino-Medina P, Fitzgerald T, Tzeng JP, McCormack LA, et al. Preferences for patient medication information: what do patients want? J Health Commun 2014;19(Suppl 2):77-88. [doi: 10.1080/10810730.2014.946114] [Medline: 25315585]
6. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the 'big five' chronic conditions: scoping review. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
7. Kelly CM, van Eaton EG, Russo JE, Kelly VC, Jurkovich GJ, Darnell DA, et al. Technology use, preferences, and capacity in injured patients at risk for posttraumatic stress disorder. Psychiatry 2017;80(3):279-285. [doi: 10.1080/00332747.2016.1271162] [Medline: 29087256]
8. Klein-Fedyshin MS. Consumer Health Informatics--integrating patients, providers, and professionals online. Med Ref Serv Q 2002;21(3):35-50. [doi: 10.1300/J115v21n03_03] [Medline: 12238015]
9. Yuan S, Hussain SA, Hales KD, Cotten SR. What do they like? Communication preferences and patterns of older adults in the United States: The role of technology. Educ Gerontol 2015;42(3):163-174. [doi: 10.1080/03601277.2015.1083392]
10. The Agency for Healthcare Research and Quality. Six Domains of Health Care Quality URL: https://www.ahrq.gov/talkingquality/measures/six-domains.html [accessed 2019-01-23] [WebCite Cache ID 75e7ML7SG]
11. Blackstone SW, Pressman H. Patient Communication in Health Care Settings: new Opportunities for Augmentative and Alternative Communication. Augment Altern Commun 2016;32(1):69-79. [doi: 10.3109/07434618.2015.1125947] [Medline: 26694249]
12. McLaughlin BK. Identifying methods to communicate with patients and enhance patient satisfaction to improve return on investment. J Manag Mark Healthc 2009;2(4):427-441. [doi: 10.1179/mmh.2009.2.4.427]
13. Mulder BC, van Lelyveld MA, Vervoort SC, Lokhorst AM, van Woerkum CM, Prins JM, et al. Communication between HIV patients and their providers: a qualitative preference match analysis. Health Commun 2016;31(1):35-46. [doi: 10.1080/10410236.2014.933017] [Medline: 25412175]
14. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 4;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
15. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]

16. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010 Jun 24;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]
17. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
18. Tistad M, Lundell S, Wiklund M, Nyberg A, Holmner Å, Wadell K. Usefulness and relevance of an eHealth tool in supporting the self-management of chronic obstructive pulmonary disease: explorative qualitative study of a cocreative process. JMIR Hum Factors 2018 Oct 26;5(4):e10801 [FREE Full text] [doi: 10.2196/10801] [Medline: 30368440]
19. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
20. Ash A, Ellis RP, Pope GC, Ayanian JZ, Bates DW, Burstin H, et al. Using diagnoses to describe populations and predict costs. Health Care Financ Rev 2000;21(3):7-28 [FREE Full text] [Medline: 11481769]
21. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
22. Betts LR, Hill R, Gardner SE. 'There's not enough knowledge out there': examining older adults' perceptions of digital technology use and digital inclusion classes. J Appl Gerontol 2019 Aug;38(8):1147-1166. [doi: 10.1177/0733464817737621] [Medline: 29165038]
23. Damnée S, Djabelkhir-Jemmi L, Ware C, Benovici J, Kérhervé H, Lewis M, et al. Exploring the impact of a group-based tablet-PC training program in older adults. Geriatr Psychol Neuropsychiatr Vieil 2019 Sep 1;17(3):336-342. [doi: 10.1684/pnv.2019.0802] [Medline: 31251216]
24. Mullangi S, Kaushal R, Ibrahim SA. Equity in the age of health care information technology and innovation: addressing the digital divide. Med Care 2019 Jun;57(Suppl 6 Suppl 2):S106-S107. [doi: 10.1097/MLR.0000000000001033] [Medline: 31095047]
25. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):pii: E645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
26. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the digital divide: age and race disparities in use of an inpatient portal. Telemed J E Health 2019 Jul 9:- [Epub ahead of print]. [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
27. Saied A, Sherry SJ, Castricone DJ, Perry KM, Katz SC, Somasundar P. Age-related trends in utilization of the internet and electronic communication devices for coordination of cancer care in elderly patients. J Geriatr Oncol 2014 Apr;5(2):185-189. [doi: 10.1016/j.jgo.2013.11.001] [Medline: 24495698]
28. Samiei V, Puteh SE, Manaf MR, Latip KA, Ismail A. Are Malaysian diabetic patients ready to use the new generation of health care service delivery? A telehealth interest assessment. Malays J Med Sci 2016 Mar;23(2):44-52 [FREE Full text] [Medline: 27547114]
29. Duplaga M. A cross-sectional study assessing determinants of the attitude to the introduction of eHealth services among patients suffering from chronic conditions. BMC Med Inform Decis Mak 2015 Apr 19;15:33 [FREE Full text] [doi: 10.1186/s12911-015-0157-3] [Medline: 25927312]
30. Nelson LA, Mulvaney SA, Gebretsadik T, Ho Y, Johnson KB, Osborn CY. Disparities in the use of a mHealth medication adherence promotion intervention for low-income adults with type 2 diabetes. J Am Med Inform Assoc 2016 Jan;23(1):12-18 [FREE Full text] [doi: 10.1093/jamia/ocv082] [Medline: 26186935]
31. Drewes C, Kirkovits T, Schiltz D, Schinkoethe T, Haidinger R, Goldmann-Posch U, et al. eHealth acceptance and new media preferences for therapy assistance among breast cancer patients. JMIR Cancer 2016 Sep 14;2(2):e13 [FREE Full text] [doi: 10.2196/cancer.5711] [Medline: 28410189]
32. Hofstede J, de Bie J, van Wijngaarden B, Heijmans M. Knowledge, use and attitude toward eHealth among patients with chronic lung diseases. Int J Med Inform 2014 Dec;83(12):967-974. [doi: 10.1016/j.ijmedinf.2014.08.011] [Medline: 25269992]
33. Anglada-Martínez H, Rovira-Illamola M, Martin-Conde M, Sotoca-Momblona JM, Codina-Jané C. mHealth intervention to improve medication management in chronically ill patients: analysis of the recruitment process. Postgrad Med 2016 May;128(4):427-431. [doi: 10.1080/00325481.2016.1170580] [Medline: 27007065]
34. Smith R, Menon J, Rajeev JG, Feinberg L, Kumar RK, Banerjee A. Potential for the use of mHealth in the management of cardiovascular disease in Kerala: a qualitative study. BMJ Open 2015 Nov 17;5(11):e009367 [FREE Full text] [doi: 10.1136/bmjopen-2015-009367] [Medline: 26576813]
35. Kamis K, Janevic MR, Marinec N, Jantz R, Valverde H, Piette JD. A study of mobile phone use among patients with noncommunicable diseases in La Paz, Bolivia: implications for mHealth research and development. Global Health 2015 Jul 4;11:30 [FREE Full text] [doi: 10.1186/s12992-015-0115-y] [Medline: 26141528]

36. Hargittai E. Second-level digital divide: differences in people's online skills. First Monday 2002;7(4):2002 ISSN-1396-0466 [FREE Full text] [doi: 10.5210/fm.v7i4.942]
37. van Deursen AJ, Helsper EJ. A nuanced understanding of Internet use and non-use among the elderly. Eur J Commun 2015;30(2):171-187. [doi: 10.1177/0267323115578059]
38. Wei K, Teo H, Chan HC, Tan BC. Conceptualizing and testing a social cognitive model of the digital divide. Inf Syst Res 2011;22(1):170-187. [doi: 10.1287/isre.1090.0273]
39. van Deursen AJ, van Dijk JA. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media Soc 2019 Feb;21(2):354-375 [FREE Full text] [doi: 10.1177/1461444818797082] [Medline: 30886536]
40. Mitchell U, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 9;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
41. Kumar D, Hemmige V, Kallen MA, Giordano TP, Arya M. Mobile phones may not bridge the digital divide: a look at mobile phone literacy in an underserved patient population. Cureus 2019 Feb 20;11(2):e4104 [FREE Full text] [doi: 10.7759/cureus.4104] [Medline: 31057998]
42. Greenberg-Worisek AJ, Kurani S, Rutten LJ, Blake KD, Moser RP, Hesse BW. Tracking healthy people 2020 internet, broadband, and mobile device access goals: an update using data from The Health Information National Trends Survey. J Med Internet Res 2019 Jun 24;21(6):e13300 [FREE Full text] [doi: 10.2196/13300] [Medline: 31237238]
43. Fang M, Canham SL, Battersby L, Sixsmith J, Wada M, Sixsmith A. Exploring privilege in the digital divide: implications for theory, policy, and practice. Gerontologist 2019 Jan 9;59(1):e1-15. [doi: 10.1093/geront/gny037] [Medline: 29750241]
44. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California Health Interview Survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
45. Ball C, Francis J, Huang K, Kadylak T, Cotten SR, Rikard RV. The physical-digital divide: exploring the social gap between digital natives and physical natives. J Appl Gerontol 2019 Aug;38(8):1167-1184. [doi: 10.1177/0733464817732518] [Medline: 29165009]
46. Schneider BC, Schröder J, Berger T, Hohagen F, Meyer B, Späth C, et al. Bridging the 'digital divide': A comparison of use and effectiveness of an online intervention for depression between Baby Boomers and Millennials. J Affect Disord 2018 Aug 15;236:243-251. [doi: 10.1016/j.jad.2018.04.101] [Medline: 29751239]
47. Graetz I, Huang J, Brand RJ, Hsu J, Yamin CK, Reed ME. Bridging the digital divide: mobile access to personal health records among patients with diabetes. Am J Manag Care 2018 Jan;24(1):43-48 [FREE Full text] [Medline: 29350505]
48. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the internet as a resource for health information: Results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
49. Delello JA, McWhorter RR. Reducing the digital divide: connecting older adults to iPad technology. J Appl Gerontol 2017 Jan;36(1):3-28. [doi: 10.1177/0733464815589985] [Medline: 26084479]
50. Bickmore TW, Paasche-Orlow MK. The role of information technology in health literacy research. J Health Commun 2012;17(Suppl 3):23-29. [doi: 10.1080/10810730.2012.712626] [Medline: 23030559]
51. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. BMC Health Serv Res 2018 Mar 27;18(1):220 [FREE Full text] [doi: 10.1186/s12913-018-2986-0] [Medline: 29587721]
52. Dholakia R. Technology and Consumption: Understanding Consumer Choices and Behaviors. New York, NY, USA: Springer-Verlag New York; 2012.
53. HIMSS Analytics. Stage 7 Case in Point: Nebraska Medicine URL: https://www.himssanalytics.org/news/stage-7-case-point-nebraska-medicine [accessed 2019-01-23] [WebCite Cache ID 75e8FPltx]
54. HIMSS Analytics. Electronic Medical Record Adoption Model (EMRAM)SM URL: https://www.himssanalytics.org/emram [accessed 2019-01-23] [WebCite Cache ID 75e7xNzQX]
55. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. Circulation 2017 Mar 7;135(10):e146-e603 [FREE Full text] [doi: 10.1161/CIR.0000000000000485] [Medline: 28122885]
56. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008 May;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]
57. Compeau DR, Higgins CA. Computer self-efficacy: development of a measure and initial test. Manag Inf Syst Q 1995;19(2):189-211. [doi: 10.2307/249688]
58. Hibbard J, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]

59. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, Moffet HH, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study. J Am Med Inform Assoc 2013;20(6):1128-1131 [FREE Full text] [doi: 10.1136/amiajnl-2012-001567] [Medline: 23676243]
60. Agency for Healthcare Research and Quality. Rockville, MD; 2013. Making Health Care Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices URL: https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/evidence-based-reports/services/quality/patientsftyupdate/ptsafetyII-full.pdf [accessed 2019-11-13]
61. Kwan JL, Lo L, Sampson M, Shojania KG. Medication reconciliation during transitions of care as a patient safety strategy: a systematic review. Ann Intern Med 2013 Mar 5;158(5 Pt 2):397-403. [doi: 10.7326/0003-4819-158-5-201303051-00006] [Medline: 23460096]
62. Vogelsmeier A, Pepper GA, Oderda L, Weir C. Medication reconciliation: A qualitative analysis of clinicians' perceptions. Res Social Adm Pharm 2013;9(4):419-430. [doi: 10.1016/j.sapharm.2012.08.002] [Medline: 23089295]
63. Anderson M. Pew Research Center. Washington, DC, USA; 2015 Oct 29. Technology Device Ownership: 2015 URL: https://www.pewresearch.org/internet/2015/10/29/technology-device-ownership-2015/ [accessed 2018-10-04]
64. Zallman L, Bearse A, West C, Bor D, McCormick D. Patient preferences and access to text messaging for health care reminders in a safety-net setting. Inform Health Soc Care 2017 Jan;42(1):32-42. [doi: 10.3109/17538157.2015.1113177] [Medline: 26864932]
65. Jayroe TJ, Wolfram D. Internet searching, tablet technology and older adults. Proc Am Soc Inf Sci Tech 2012;49(1):1-3. [doi: 10.1002/meet.14504901236]
66. Perrin A, Duggan M. Pew Research Center. Washington, DC, USA; 2015 Jun 26. Americans' Internet Access: 2000-2015 URL: https://www.pewresearch.org/internet/2015/06/26/americans-internet-access-2000-2015/ [accessed 2018-10-03]
67. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of technology: frequency of use for younger and older adults. Ageing Int 2011 Mar;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
68. Ishikawa H, Roter DL, Yamazaki Y, Takayama T. Physician-elderly patient-companion communication and roles of companions in Japanese geriatric encounters. Soc Sci Med 2005 May;60(10):2307-2320. [doi: 10.1016/j.socscimed.2004.08.071] [Medline: 15748678]
69. Schulz R, Eden J. Family caregiving roles and impacts. In: Families Caring for an Aging America. Washington, DC, USA: National Academies Press (US); 2016.
70. Harris M. Recognizing the important role of family caregivers. Home Healthc Now 2017;35(10):579. [doi: 10.1097/NHH.0000000000000623] [Medline: 29095344]

Abbreviations

AVS: after-visit summaries
CMS: Centers for Medicare and Medicaid Services
eHealth: electronic health
EHR: electronic health record
HIT: health information technology
PHR: personal health record
UNMC: University of Nebraska Medical Center

Edited by G Eysenbach; submitted 24.01.19; peer-reviewed by N Davies, T Loda, B Kane, C Jacob; comments to author 28.04.19; revised version received 22.11.19; accepted 26.01.20; published 01.06.20.

Please cite as:
Clarke MA, Fruhling AL, Sitorius M, Windle TA, Bernard TL, Windle JR
Impact of Age on Patients' Communication and Technology Preferences in the Era of Meaningful Use: Mixed Methods Study
J Med Internet Res 2020;22(6):e13470
URL: https://www.jmir.org/2020/6/e13470
doi:10.2196/13470
PMID:32478658

Review

Patient Work and Their Contexts: Scoping Review

Kathleen Yin[1], BPharm, PhD; Joshua Jung[1], BMEDSC; Enrico Coiera[1], MBBS, PhD; Liliana Laranjo[1], MD, PhD; Ann Blandford[2], PhD; Adeel Khoja[3], MBBS, MScPH; Wan-Tien Tai[1], BSc, MA; Daniel Psillakis Phillips[1], BMEDSC; Annie Y S Lau[1,2], BE, PhD

[1] Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia

[2] UCL Interaction Centre, University College London, London, United Kingdom

[3] Department of Medicine, Aga Khan University, Karachi, Pakistan

Corresponding Author:
Kathleen Yin, BPharm, PhD
Centre for Health Informatics
Australian Institute of Health Innovation
Macquarie University
Level, 75 Talavera Rd
North Ryde, 2113
Australia
Phone: 61 2 9850 2477
Email: kathleen.yin@mq.edu.au

Abstract

Background: Having patients self-manage their health conditions is a widely promoted concept, but many patients struggle to practice it effectively. Moreover, few studies have analyzed the nature of work required from patients and how such work fits into the context of their daily life.

Objective: This study aimed to review the characteristics of patient work in adult patients. Patient work refers to tasks that health conditions impose on patients (eg, taking medications) within a system of contextual factors.

Methods: A systematic scoping review was conducted using narrative synthesis. Data were extracted from PubMed, Excerpta Medica database (EMBASE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), and PsycINFO, including studies from August 2013 to August 2018. The included studies focused on adult patients and assessed one or more of the following: (1) physical health–related tasks, (2) cognitive health–related tasks, or (3) contextual factors affecting these tasks. Tasks were categorized according to the themes that emerged: (1) if the task is always visible to others or can be cognitive, (2) if the task must be conducted collaboratively or can be conducted alone, and (3) if the task was done with the purpose of creating resources. Contextual factors were grouped according to the level at which they exert influence (micro, meso, or macro) and where they fit in the patient work system (the macroergonomic layer of physical, social, and organizational factors; the mesoergonomic layer of household and community; and the microergonomic triad of person-task-tools).

Results: In total, 67 publications were included, with 58 original research articles and 9 review articles. A variety of patient work tasks were observed, ranging from physical and tangible tasks (such as taking medications and visiting health care professionals) to psychological and social tasks (such as creating coping strategies). Patient work was affected by a range of contextual factors on the micro, meso, or macro levels. Our results indicate that most patient work was done alone, in private, and often imposing cognitive burden with low amounts of support.

Conclusions: This review sought to provide insight into the work burden of health management from a patient perspective and how patient context influences such work. For many patients, health-related work is ever present, invisible, and overwhelming. When researchers and clinicians design and implement patient-facing interventions, it is important to understand how the extra work impacts one’s internal state and coping strategy, how such work fits into daily routines, and if these changes could be maintained in the long term.

(J Med Internet Res 2020;22(6):e16656) doi:10.2196/16656

KEYWORDS

self-care; burden of illness; self-management

Introduction

Background

Chronic diseases are increasingly prevalent as the world's population ages, requiring millions of patients to adjust their lifestyle and manage their health [1]. However, patient work and patient ergonomics, defined, respectively, as the combination of all health-related tasks and the contextual factors influencing the tasks [2], are given little attention. Existing studies have investigated the influences of individual barriers for health management, such as time requirements [3], the burden of comorbidity [4,5], or complexities in specific diseases [6-8]. Few studies have analyzed the character of patient work and how such work fits into the context of an individual's life.

Patients conduct a variety of cognitive, visible, and collaborative work to accommodate health conditions and treatments, with such tasks changing throughout the *illness journey* as treatments are introduced or removed [9-13]. Using a work ergonomic system, Holden et al [2,14] posited that *patient work* is affected by a *patient work system*, which incorporates contextual factors affecting the performance of work [15-20]. A poor integration of patient work into the existing context and routine can generate excess stress, potentially contributing to noncompliance and suboptimal health outcomes [21,22]. The patient work system [2] groups all contextual factors into 3 levels: a microergonomic level (further separated into people, task, and tools), a mesoergonomic level including household and community, and a macroergonomic level (physical, social, and organizational), separating the influences that are inherent within the task or the patient from those that came from daily life.

Although health-related tasks are included in the umbrella terms of *self-management* (defined as actions and processes that people with a health problem intentionally perform to manage health in partnership with health care professionals [23]) and *self-care* (a more loose definition encompassing all things people do to manage and improve their health [24]), such tasks have not been reviewed to reflect how they relate to different aspects of the patient's life. Similarly, there have been few attempts to assess how patient work is influenced by different contextual factors. In particular, although these tasks and factors have been explored in specific conditions [2,25], a review across all health conditions remains lacking. Therefore, a scoping review was chosen to rapidly gain an understanding of this nascent field, following guidance from Arksey and O'Malley [26], who noted that the method was appropriate "especially where an area is complex or has not been reviewed comprehensively before."

Objectives

In this scoping review, we review the existing literature on patient work, examining what people do in self-management and why they do or do not undertake certain tasks [2]. We also examine the different levels of patient work contexts, where digital interventions could play a supporting role. Specifically, digital technologies can offer health advice based on the immediate tasks and contexts around the patient [27], making digital apps a viable solution to supporting patient work through one's life.

Methods

Search Strategy

The search was conducted using a modified participants, interventions, comparisons, and outcome strategy, which stated that the research question for a review must include the population, intervention, comparison, and outcome. Our research question was "What are the characteristics of patient work in adult patients?," with the population being adult patients, intervention being the presence of a health condition, comparison being daily life before diagnosis, and the outcome being the characteristics of patient work. A search was conducted on August 23, 2018, in PubMed, Excerpta Medica database (EMBASE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), and PsycINFO, including all articles published from August 2013 onward. The search terms were designed to capture publications that depicted work conducted by patients to maintain their health and how such work was limited or facilitated by contextual factors. Only articles published in English were included in the search. Multimedia Appendix 1 provides the complete search strategy.

Inclusion and Exclusion Criteria

Articles were eligible if they focused on adult patients, included a qualitative component, and focused on assessing the impact of (1) physical tasks that patients undertake to manage their health, (2) cognitive tasks associated with managing health, *or* (3) contextual factors facilitating or restricting the physical or cognitive tasks.

Articles were excluded if they only focused on nonpatients (eg, caregivers); if they only presented biological or physiological data; if they only addressed health-related contextual factors that do not affect the work patients do (eg, computing systems in a hospital); were opinion articles or protocol papers; *or* if they focused on the design or evaluation of a measuring instrument, tool, or intervention.

Study Screening and Data Extraction

Multimedia Appendix 1 provides details on abstract and full-text screening as well as data extraction from full-text articles. Each abstract and full-text was screened independently by 2 researchers, and each full-text article was also screened independently by 2 researchers. The interrater score for abstract screening was 0.39 (fair agreement) [28] and for full-paper screening was 0.30 (fair agreement) [28]. Disagreements were resolved by having a third independent reviewer review the conflicting article and make a final decision. Data extraction was conducted by 3 researchers who met regularly to address concerns and to ensure that data extraction was conducted consistently.

Analysis Framework

We conducted a narrative synthesis on the patient work tasks and extracted contextual factors. The patient work tasks were assessed using the patient work model initially proposed by Corbin and Strauss [29,30], which indicates patient work as *illness work* that is influenced by *everyday life work* and *biographical work*. The work results in 3 types of tasks for self-management: medical management (eg, planning doctor appointments), emotional management (eg, dealing with anxiety and fear), and role management (eg, balancing one's role in the family with one's illness). We consulted a modified version of Corbin and Strauss's model by Dack et al [31], which encompassed the 3 types of tasks for self-management. For contexts, we used the patient work system proposed by Holden et al [2,14]. The patient work system encompasses a microergonomic aspect, which describes how contextual factors from the patient, task, and tools involved [14,32] affect the work done; a mesoergonomic layer of household and community; and a macroergonomic aspect where physical, social, and organizational influences [2,14] are outlined.

Work tasks were identified from the included articles, were clustered based on common themes that emerged, and were consulted against Dack et al's model [31]. Contextual factors were grouped according to where they fit within the dimensions outlined in the patient work framework (patient, tasks, tools, physical, social, and organizational) [2,14,32] and the level at which they exert influence (micro, meso, or macro) [33,34]. We identified microlevel contextual factors as aspects that only affected the patient's body or were psychological factors. Mesolevel contextual factors were the influences of people socially close to the patient or influences exerted by the immediate physical surroundings or social circumstances (eg, finances) of the patient [35]. We identified macrolevel contextual factors as aspects that arose from the society, culture, or geopolitical entity in which the patient lived. The micro-meso-macro framework has been used extensively in health care, ranging from describing the quality of life assessments at different levels of decision making [36], patient-reported outcome measures in hospital palliative care [34], enablers and barriers affecting nursing practices [37], policy implementation of health leadership [38], to health care priority setting across different countries [39].

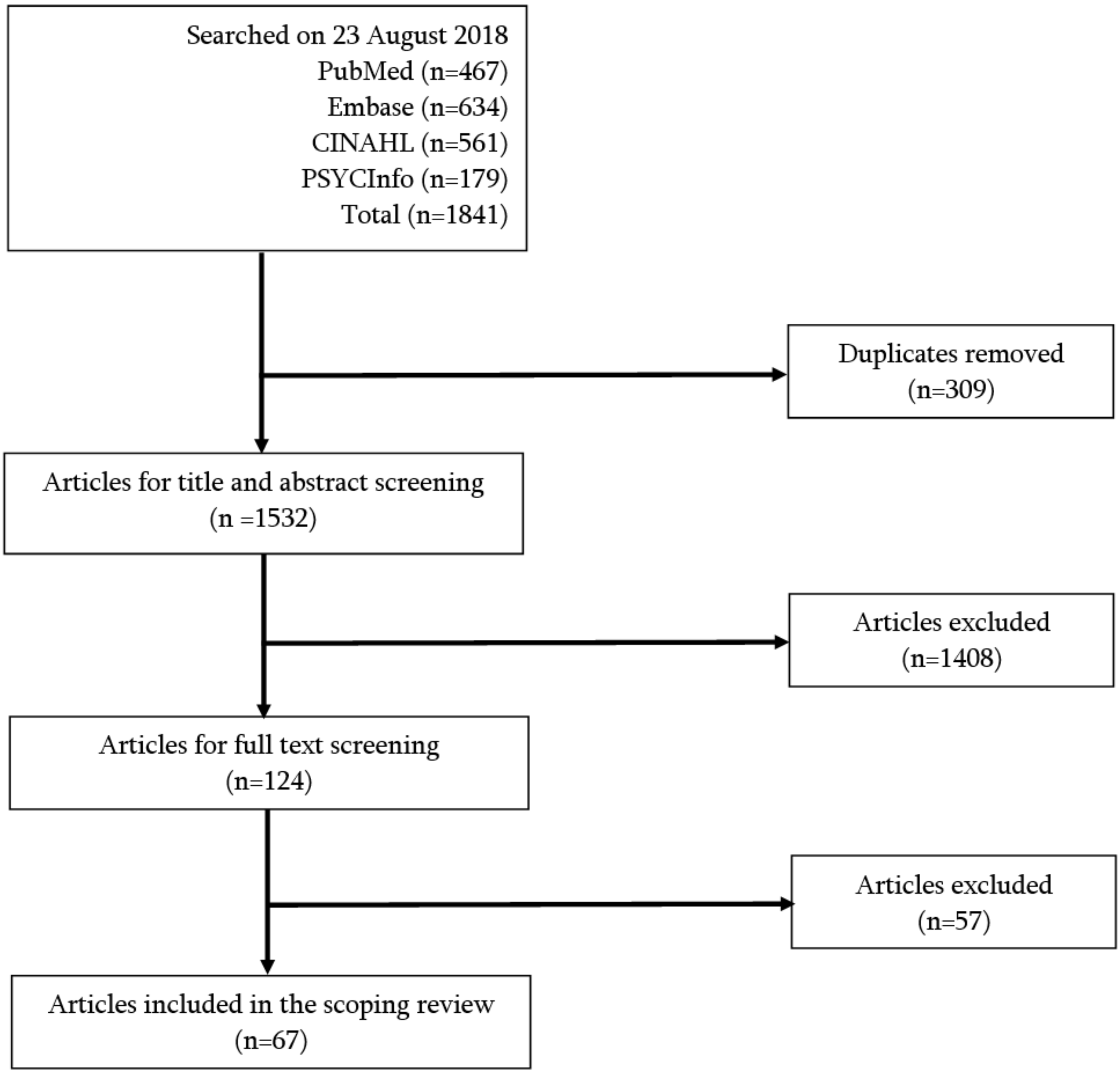
The articles were initially read by KY, JJ, and DP to extract passages that described patient work tasks or contextual influences on patient work. The 3 researchers then compiled our findings to reach a consensus and held frequent meetings over 1 month to sort all identified patient work tasks and contextual influences into themes. Reviewers decided by consensus to categorize the work tasks and the contextual factors based on the commonalities observed.

Results

Screening Process

Figure 1 outlines the screening process. Multimedia Appendix 2 gives more details about the screening process.

Figure 1. Number of articles included at each stage of the screening process.



Characteristics of Included Articles

Among the 67 articles included in the scoping review, 58 were original research and 9 were reviews. Semistructured interviews were the most common research strategy employed, and the articles studied 37 different health conditions. Multimedia Appendices 3 and 4 provide more information on the type and characteristics of the included articles.

Patient Work Tasks Conducted by Participants

Patient work occurs in 2 forms: visible and cognitive work. Visible work is performed within a physical space (such as driving to visit doctors). Such tangible tasks could be observed by other people and are easy for others to intervene. Cognitive work, on the other hand, is completely unseen by others (eg, mentally counting calories throughout the day). Such tasks would not be revealed unless the patient discussed the information directly, and the task could remain hidden even from family and close friends.

Patient work tasks can also be conducted collaboratively or alone. Some tasks must be conducted in collaboration with others (eg, visiting health professionals), whereas cognitive tasks are always conducted alone (eg, developing mental coping strategies). Most patient work tasks, however, existed between these 2 extremes and could be conducted collaboratively or alone, depending on the contextual influences.

All patient work tasks also consumed resources such as time [40,41], physical energy [42], or social support. However, although many patient work tasks consumed resources to maintain one’s state of health, some tasks had the precise purpose of creating more resources, such as learning about one’s health condition or attending patient support group meetings. The prevalence of resources available to a patient serves as a buffer to mitigate the effect of sudden changes (eg, having enough funds to undergo surgery), and tasks that increase resources can ultimately increase the patient’s capacity to cope [43].

In total, 6 different categories of patient work tasks emerged along the 3 aforementioned axes and are listed in Textbox 1. The categories were distinctive based on whether (1) the task was always visible or not, (2) the task must be conducted collaboratively or not, and (3) the purpose of the task was to create resources or not.

Tasks were further divided into 21 specific types of tasks according to the reviewers' consensus. Multimedia Appendix 5 provides a list of tasks identified in each included article, and Multimedia Appendix 6 provides a thorough description of each task type together with examples.

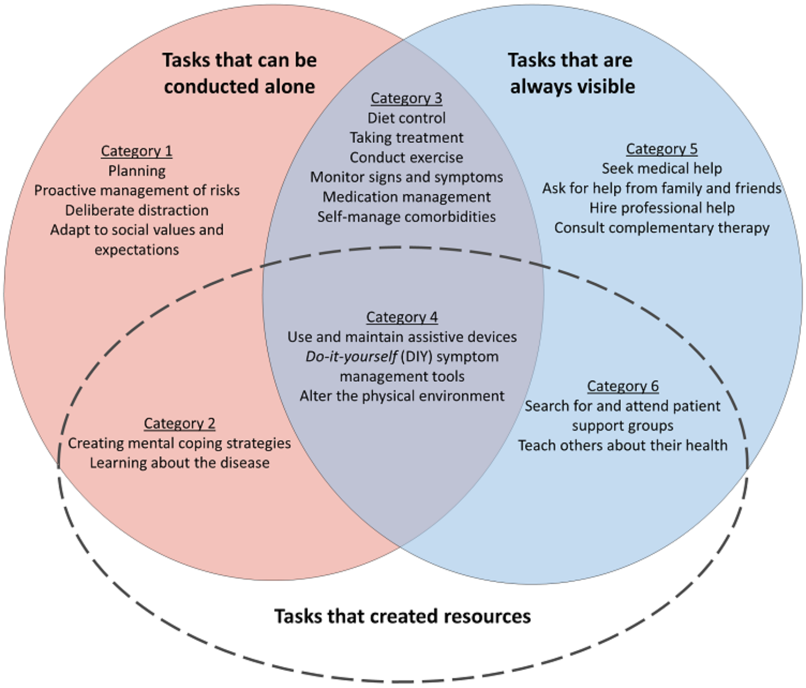
Figure 2 demonstrates how the tasks fit into the 6 identified categories and the relationship of the categories with each other.

Textbox 1. Classification of patient work tasks.

Category and description of the patient work tasks

- Category 1: Not always visible, can be conducted alone, and consumed resources
- Category 2: Not always visible, can be conducted alone, and purposefully created resources
- Category 3: Visible, can be conducted alone, and consumed resources
- Category 4: Visible, can be conducted alone, and purposefully created resources
- Category 5: Visible, must be conducted collaboratively, and consumed resources
- Category 6: Visible, must be conducted collaboratively, and purposefully created resources

Figure 2. Patient work tasks were categorized according to their alignment along the 3 axes of collaboration, visibility, and creating resources.



Category 1: Tasks That Are Not Always Visible, Can Be Conducted Alone, and Consumed Resources

Category 1 tasks made up a large portion of the instances of patient work tasks identified in the included studies, with the most prevalent tasks being *planning* [18,20,40,44-57] and *proactive management of risks* [20,40-42,45-47,49,50,52,57-78]. Tasks in category 1 are characterized by their pervasiveness and volume. Patients are constantly thinking about their prospective health-related plans and mitigating health risks, even when they are not currently experiencing symptoms or remaining in a health care setting. Significantly, health-related work in this category includes tasks that do not necessarily improve health outcomes (such as *adapt to social values and expectations* [19,20,48,55,57,58,61,64,70,79-82]). Despite the volume of work in this category, patients do not often discuss such tasks with others, making such work unacknowledged and sometimes taken for granted by the patient and their families, receiving very little organizational or psychological support.

Category 2: Tasks That Are Not Always Visible, Can Be Conducted Alone, and Purposefully Created Resources

Work tasks included in category 2 are not so much about managing one's symptoms, but work that enables patients to manage their symptoms better in the future. *Create mental coping strategies* [14,18,20,40,43,45,46,48-51,61,75-80,83-89] included overcoming emotional barriers [18], drawing upon spiritual beliefs [78], eventually coming to terms with a *new* *body* and a *new normal* [14,20,87,89] that enables patients to proactively engage with treatment. Patients also looked for relevant knowledge, either from written literature or by asking other people, in work tasks under *learning about the disease* [16,20,42,48,57,59,60,64,67,73,74,78,80,81,84,90]. Intellectual endeavors to understand the implications of symptoms [52] are also included here.

Category 3: Tasks That Are Visible, Can Be Conducted Alone, and Consumed Resources

Category 3 tasks included self-management tasks such as *diet control* [14,42,48,49,52,53,57,59,63-68,71,75,77-79,82,85,87,90,91], *conduct exercise* [19,20,25,40-42,45,49,56,59,63,66,67,73,77,79,82,87,90,91], *taking treatment* [2,18,42,47,49,57,63,64,66-69,71,73-75,78,82,84,85,90-94], *monitor signs and symptoms* [2,14,41,42,44,45,57,65,67-69,71,85,87,90,92,95], *medication management* [40,43,44,51,52,57,59,71,75,84,90,95,96], and *self-manage comorbidities* [14,41,44,59,75,79,82,97]. The effective carrying out of tasks in this category is underpinned by category 1 and 2 tasks, benefiting from good planning and attitude changes. Although patients recognize this category to be a significant drain on their resources (whether in terms of time [40,41], physical capacity [42], or appropriate knowledge [19,47,71]), they are aware of the importance of such tasks and noted that support for such tasks is already in place from public health initiatives and health professionals.

Category 4: Tasks That Are Visible, Can Be Conducted Alone, and Purposefully Created Resources

Category 4 was the only category that explicitly described how patients changed their physical environment and tools. Tasks in this category (eg, *use and maintain assistive devices* [16,20,42,50,56,58,59,62,67,69,72,74,89,90,95,97,98], *do-it-yourself symptom management tools* [18,19,47,59,62,66,74,77,95], and *alter the physical environment* [46,62,93]) had an immediate physical return in the form of better tools or a more comfortable physical environment, with the tools ranging from day-to-day items (eg, shoe insoles [97]) to specialized equipment (such as home oxygen tanks [59]). Patients are very aware of how this category is an investment to improve the quality of life as well as what kind of support (mainly financial) is available from health organizations.

Category 5: Tasks That Are Visible, Must Be Conducted Collaboratively, and Consumed Resources

This category included tasks where patients sought help from other people, including *seek medical help* [2,14,18,40,41,44,49-52,58,65,75,79,80,83,89,93,96,97], *ask for help from family and friends* [18,40,46,47,50,51,57,60,63,67,69,73,77,93,95,98], *hire professional help* [20,43,52,69,95,98], and *consult complementary therapy* [58,59,64,66,84]. This category also included unplanned interactions, such as visits to the emergency department [41,89]. The patient reported only initiating these tasks when issues have escalated beyond their individual control [41,52,89] and seeking help became a necessity. Being the most visible tasks observable by health professionals, health care systems have traditionally paid close attention to these interactions.

Category 6: Tasks That Are Visible, Must Be Conducted Collaboratively, and Purposefully Created Resources

The last category, category 6, included tasks that contributed specifically to building new social resources for the patient, *searching for and attending patient support groups* [57,59,67,73,99], and *teaching others about their health* [2,51,57,71,89]. Patients actively create new social circles and recruit other people into their lives, whether by joining patient support groups or educating family and friends about their health [2,51,57,71,89]. In these interactions, the patient acts as either a peer or an expert, instead of the party receiving help. They proactively share their own health information and self-care strategies with others and act to support other patients along the *illness journey*.

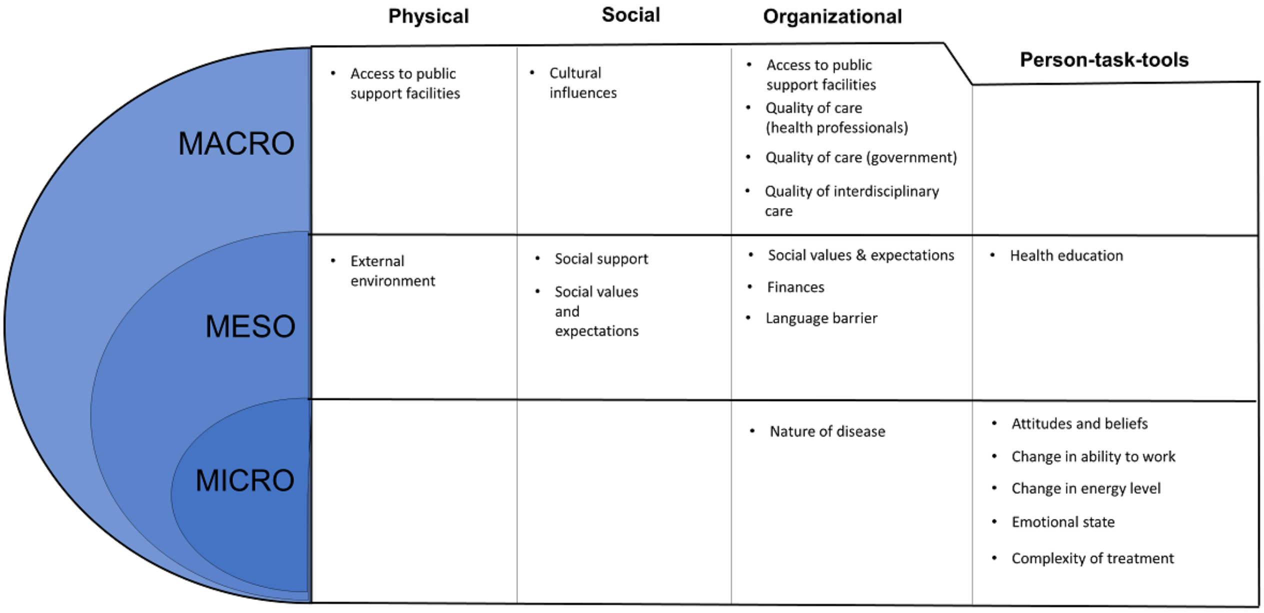
Contextual Factors Influencing Patient Work Tasks

We identified 17 types of contextual factors that influence patient work tasks. Multimedia Appendix 5 provides a list of contextual factors identified in each included article, and Multimedia Appendix 7 provides a thorough description of each contextual factor with examples. The contextual factors were grouped into how they fit within the patient work system (either in the macroergonomic triad of physical, social, and organizational factors or within the microergonomic triad of person-task-tools) [2,14,32]. As the patient work system has 3 layers, with households and communities comprising the mesoergonomic layer between the macro and microergonomic contexts, we also grouped contextual factors based on the level (macro, meso, or micro) at which they affected patient work tasks [33,34]. Although Holden et al [100] have identified *mesoergonomics* as the study of the relationship between variables in different layers of an ergonomics system, the *meso* contextual factors we identified here correspond to factors that sit between the micro- and macroergonomics levels.

Microlevel Contextual Factors

We identified microlevel contextual factors as aspects that only affected health management by influencing a patient’s body or were psychological factors that arose from the patient’s mindset. Such factors significantly influence patients on a personal level and are included in the microergonomic triad of person-task-tools within the patient work system by Holden et al [2], as indicated in Figure 3. Microlevel factors in the review were mostly psychological factors, not always noticeable to family members or health professionals [2,45,61,70]. Such factors echo the type of patient work tasks that were not always visible, indicating a concrete need for resources that assist with psychological coping and adopting self-care into one’s routine.

Figure 3. Contextual factors that influence patient work, separated on the basis of their allocation in the patient work system and whether they act on a macro, meso, or micro level.



Mesolevel Contextual Factors

Mesolevel contextual factors are aspects that affect self-management because of the influences of people socially close to the patient or influences exerted by the immediate surroundings or circumstances of the patient. These factors equally affected the physical, social, and organizational macroergonomic domains of the patient work system [2]. Contextual factors identified at this level, such as language barriers and social support, are widely recognized by health care systems, and supportive measures are often already in place [43,55,63,86]. Moreover, efforts to locally improve these contextual factors can significantly improve the patient’s health management, making such contextual factors an existing focus for intervention from health professionals.

Macrolevel Contextual Factors

Macrolevel contextual factors arise from the society, culture, or geopolitical entity in which the patient lived. These factors affected the organizational aspect of the patient work system [57,58,69,82]. Changing such contextual factors usually requires significant political or population health action, and individual patients and health care professionals often struggle to influence such factors on their own [47,57,63,69,82].

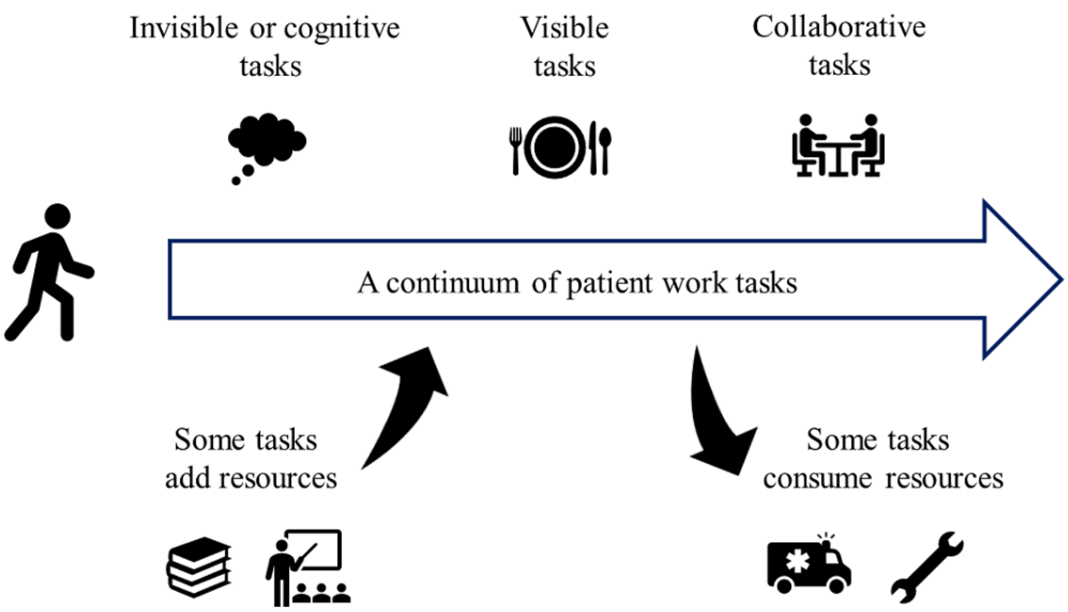
Discussion

Principal Findings

The purpose of this scoping review was to analyze the existing literature on the tasks that health conditions imposed on patients and the contextual factors affecting these tasks [2]. Our results indicate a continuum of patient work tasks, through which the patient moves from tasks that are cognitive only to tasks that are always visible and tasks that are always conducted collaboratively, while also experiencing how some tasks generated resources for the future, whereas others mainly consumed resources, as seen in Figure 4.

Contextual factors were mapped out by analyzing the factors according to their micro-, meso-, or macrolevel of influence and where they fitted along the patient work system, whether within the macroergonomic factors of physical, social, and organizational factors or the microergonomic triad of person-tasks-tool [14]. Although all contextual factors fit into the patient work system, there seems to be a *mesoergonomic* layer not specified in the patient work system that acted as a bridge between the macroergonomic influences and the more personal and psychological microergonomic contextual factors.

Figure 4. A continuum of patient work tasks that the patient moves through in their illness journey.



Limitations

Our review has some limitations. We only included articles published in English and did not have access to studies in other languages. The relationship between the different aspects of work tasks is propositional and has not been empirically tested. We also focused on articles that addressed work conducted by patients themselves, as opposed to work conducted by caregivers.

Our review also did not cover detailed physical ergonomic factors of the patient’s vicinity, such as the physical properties of the areas where patients placed their medications, as the included articles did not contain these details.

Comparison With Existing Literature

Corbin and Strauss [30] separated self-management into 3 different kinds of activities: emotional management, medical management, and role management. Category 1 (not always visible, can be conducted alone, and consumed resources) and category 2 (not always visible, can be conducted alone, and purposefully created resources) relate to cognitive workflow [101] and fit into emotional management within the Corbin and Strauss’s model. Within the cognitive workflow, which consists of sensemaking [102,103], planning, monitoring, decision making [9], and coordinating, the person makes the cognitive preparations necessary before conducting an action. This was reflected in activities such as planning and adjusting routines to *make space* for health-related work. Other cognitive tasks, such as finding information on the web and developing mental strategies, contribute to helping the patient to intellectually understand and psychologically accept their health condition.

Patients consider these tasks a part of life and rarely mention such tasks to health professionals despite its prevalence, reflecting a glaring need for more systematic support of these *invisible* tasks. Although this scoping review defined *not visible work* as work that is conducted primarily cognitively and thus *invisible* in the literal sense, tasks in categories 1 (not always visible, can be conducted alone, and consumed resources) and category 2 (not always visible, can be conducted alone, and purposefully created resources) can also fit into the definition of *invisible work* as it is used in ergonomics [104]—work that is taken for granted, underacknowledged, and undervalued. These tasks are only recognized as a burden when patients find them overwhelming.

A large body of literature exists on supporting patients with psychological distress, including psychotherapy [105] for patients with advanced cancer, cognitive behavioral therapy, and animal therapy [106], yet everyday tasks such as planning and establishing routines are much more nebulous to support. Patients reported using a variety of basic planning tools, such as notebooks and calendars with appointments written in them, and it is possible that a digital extension of such tools, such as integrated digital diaries, could better assist in this space.

Tasks in categories 3 (visible, can be conducted alone, and consumed resources) and category 4 (visible, can be conducted alone, and purposefully created resources) represent health-related work acted out by individual patients, corresponding to medical management [30]. The dimension of lifestyle changes [107] is incorporated in these tasks as patients carry out exercise, modify their diet, and sort their medications into doses, work that the clinician traditionally ascribes to self-management. It is well recognized that patients need to have sufficient resources such as time [40,41], access to affordable facilities, or physical capacities to perform tasks satisfactorily [42]. Health professionals and health organizations have focused on assisting with these tasks for many years, and many patients are aware of the presence of such support programs. Category 4 (visible, can be conducted alone, and purposefully created resources) present as a slightly different category that initially consumes time and money but ultimately results in a large increase in capacity, such as the purchase of assistive devices, which can significantly improve the efficacy of self-care in the future.

Tasks in categories 5 (visible, must be conducted collaboratively, and consumed resources) and category 6 (visible, must be conducted collaboratively, and purposefully created resources) are tasks that are inherently collaborative, bringing self-management out of the individual's personal lives and interacting with role and relationship management [30] regarding health professionals or family. These tasks also correspond to the self-care dimensions of communication [107,108] and obtaining help [109]. Some tasks in category 5 have the patient play a passive role (eg, visiting the emergency department), whereas activities in category 6 saw the patient interact with others as a peer or an expert (eg, attending patient support groups and teaching others about the disease and treatments). When a patient acts as an equal partner in a health-related relationship and feels the relationship to be a positive and empowering experience, the psychological and social benefits expand the patient's capacity [43] and social resilience [15,110]. However, although the health system noted the benefits of such tasks and encouraged collaborative and equal decision making between the patient and the clinician, category 6 tasks were the least prevalent in our review, indicating that such tasks may still be taken up by only a small section of more informed and proactive patients.

Contextual Factors in the Patient Work Framework

The patient work system [2] separates the patient's surroundings into macroergonomic and microergonomic categories, with a middle household and community levels. In the macroergonomics layer, physical contexts described influences of the physical world and social contexts described the influence of other people, whereas organizational contexts encompassed temporal organization (daily routines), societal organization (finances and family roles), and political organization (health system and legal issues). In the microergonomics layer, person described the characteristics of the people involved, task described the inherent challenges of the task, and tools described the method or tool used to perform the task.

This review further separated the contextual influences into 3 layers: macro, meso, and micro. At the macrolevel, the patient work system addressed many factors that affected patients on a governmental or cultural level. The mesolevel, which was not explicitly stated in the patient work system, appears to act as a bridge transporting macrolevel values and expectations down to the individual. For example, cultural values filtered down to social expectations, and the structure of the health care system emphasized the impact of financial problems. Although not addressed within the patient work system, the meso layer certainly exists and exerts its influence through the people and the environment closest to the patient.

On the microlevel, most of the identified contextual factors fit into the person-task-tools triad, and most of those factors were within *person*. Patients in our review reported a variety of emotional and mental influences that altered their actions, indicating their motivation and belief in their capacity was sometimes more important than the resources available to them. Patients experiencing different health conditions were affected by common emotional states, such as fear of medication side effects and dreading the future [67,69,71,75,80]. As an element of concern, patients in our scoping review rarely reported consulting mental health professionals about these fluctuating emotions, opting instead to *brave through* on their own [45,58,60,71,75,96]. Although there is literature indicating that different personality types respond differently to intervention styles [111-113], it is also probable that patients need to be aware of the available psychological help or be motivated to use such services to receive sufficient help through their psychological journey.

Conclusions

This review aimed to provide insight into the design and implementation of self-management interventions by understanding where the health-related work burden lies for community-based patients and how context influences such work. In our scoping review, we found a high prevalence of patient work tasks conducted in the privacy of the patient's own lives—often alone, cognitive or unacknowledged, and consuming resources. However, despite continuous efforts to improve community-based self-management, there are no definitive models of intervention addressing this need, and patients continue to struggle to incorporate self-management into their daily lives.

Innovative digital technologies, such as using digital devices to track and monitor one's health, may play a role in supporting individuals with these invisible and solitary health tasks. Similarly, more personalized and flexible treatment might be achieved by sociotechnical interventions, which carefully consider how interactions with the health care system could affect different facets of an individual's daily life.

Existing self-management literature, traditionally only focusing on the physical aspects of health-related work, has been enriched by unveiling the internal experiences of patients. As research increasingly considers health as an interplay between biology, psychology, and sociology, we are beginning to better assess the various layers of patient work and contexts that can influence the implementation of self-care.

Acknowledgments

The authors would like to thank Ms Yi-Jen Lu, Ms Natasha Hughes, and Mr Ivan Khizhnyak for their assistance with paper screening and data extraction. This research was supported by the National Health and Medical Research Council grant APP1134919 (Centre for Research Excellence in Digital Health). AL was supported by the New South Wales Health Early-Mid Career Fellowship.

Authors' Contributions

KY and JJ designed the study, developed the search strategy, conducted the searches, screened the retrieved papers, extracted relevant information, and drafted the paper as the first authors. EC conceptualized the study and helped with drafting the paper, and LL, AK, and WT contributed toward search strategy development, paper screening, and data extraction. AB helped with developing the search strategy and writing drafts of the paper, and DP assisted with data extraction. AL contributed throughout the project, starting from conceptualization, to study design, search strategy development, and editing subsequent drafts of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and screening processes.
[DOCX File , 14 KB - jmir_v22i6e16656_app1.docx]

Multimedia Appendix 2
Screening results.
[DOCX File , 12 KB - jmir_v22i6e16656_app2.docx]

Multimedia Appendix 3
Characteristics (original research articles).
[DOCX File , 57 KB - jmir_v22i6e16656_app3.docx]

Multimedia Appendix 4
Characteristics (reviews).
[DOCX File , 18 KB - jmir_v22i6e16656_app4.docx]

Multimedia Appendix 5
[DOCX File , 57 KB - jmir_v22i6e16656_app5.docx]

Multimedia Appendix 6
Example of patient worktasks.
[DOCX File , 18 KB - jmir_v22i6e16656_app6.docx]

Multimedia Appendix 7
Examples of contextual factors.
[DOCX File , 16 KB - jmir_v22i6e16656_app7.docx]

References

1. World Health Organisation. Global Strategy and Action Plan on Ageing and Health. Geneva, Switzerland: World Health Organisation; 2017.
2. Holden RJ, Valdez RS, Schubert CC, Thompson MJ, Hundt AS. Macroergonomic factors in the patient work system: examining the context of patients with chronic illness. Ergonomics 2017 Jan;60(1):26-43 [FREE Full text] [doi: 10.1080/00140139.2016.1168529] [Medline: 27164171]
3. Yen L, McRae I, Jowsey T, Gillespie J, Dugdale P, Banfield M, et al. Health work by older people with chronic illness: how much time does it take? Chronic Illn 2013 Dec;9(4):268-282. [doi: 10.1177/1742395313476720] [Medline: 23463322]
4. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician 2014 Dec;60(12):1123-1133 [FREE Full text] [Medline: 25642490]
5. Koch G, Wakefield BJ, Wakefield DS. Barriers and facilitators to managing multiple chronic conditions: a systematic literature review. West J Nurs Res 2015 Apr;37(4):498-516. [doi: 10.1177/0193945914549058] [Medline: 25193613]
6. Douglass J, Graves P, Gordon S. Self-care for management of secondary lymphedema: a systematic review. PLoS Negl Trop Dis 2016 Jun;10(6):e0004740 [FREE Full text] [doi: 10.1371/journal.pntd.0004740] [Medline: 27275844]
7. Jaarsma T, Cameron J, Riegel B, Stromberg A. Factors related to self-care in heart failure patients according to the middle-range theory of self-care of chronic illness: a literature update. Curr Heart Fail Rep 2017 Apr;14(2):71-77 [FREE Full text] [doi: 10.1007/s11897-017-0324-1] [Medline: 28213768]

8. Jolly K, Majothi S, Sitch AJ, Heneghan NR, Riley RD, Moore DJ, et al. Self-management of health care behaviors for COPD: a systematic review and meta-analysis. Int J Chron Obstruct Pulmon Dis 2016;11:305-326 [FREE Full text] [doi: 10.2147/COPD.S90812] [Medline: 26937183]
9. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
10. Kaptein A, Fischer M, Scharloo M. Self-management in patients with COPD: theoretical context, content, outcomes, and integration into clinical care. Int J Chron Obstruct Pulmon Dis 2014;9:907-917 [FREE Full text] [doi: 10.2147/COPD.S49622] [Medline: 25214777]
11. Apps LD, Harrison S, Williams JE, Hudson N, Steiner M, Morgan M, et al. How do informal self-care strategies evolve among patients with chronic obstructive pulmonary disease managed in primary care? A qualitative study. Int J Chron Obstruct Pulmon Dis 2014;9:257-263 [FREE Full text] [doi: 10.2147/COPD.S52691] [Medline: 24600218]
12. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
13. Clark NM, Gong M, Kaciroti N. A model of self-regulation for control of chronic disease. Health Educ Behav 2001 Dec;28(6):769-782. [doi: 10.1177/109019810102800608] [Medline: 11720277]
14. Holden RJ, Schubert CC, Mickelson RS. The patient work system: an analysis of self-care performance barriers among elderly heart failure patients and their informal caregivers. Appl Ergon 2015 Mar;47:133-150 [FREE Full text] [doi: 10.1016/j.apergo.2014.09.009] [Medline: 25479983]
15. Lecerof S, Stafström M, Westerling R, Östergren PO. Does social capital protect mental health among migrants in Sweden? Health Promot Int 2016 Sep;31(3):644-652. [doi: 10.1093/heapro/dav048] [Medline: 26048867]
16. Blandford A, Farrington K, Mayer A, Walker D, Rajkomar A. Coping strategies when self-managing care on home haemodialysis. J Ren Care 2015 Sep 2;7(5):222-228. [doi: 10.12968/jorn.2015.7.5.222]
17. Burgess E, Reddy M, Davenport A, Laboi P, Blandford A. 'Tricky to Get Your Head Around': Information Work of People Managing Chronic Kidney Disease in the UK. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300895]
18. Flurey CA, Morris M, Richards P, Hughes R, Hewlett S. It's like a juggling act: rheumatoid arthritis patient perspectives on daily life and flare while on current treatment regimes. Rheumatology (Oxford) 2014 Apr;53(4):696-703 [FREE Full text] [doi: 10.1093/rheumatology/ket416] [Medline: 24357813]
19. Halls S, Dures E, Kirwan J, Pollock J, Baker G, Edmunds A, et al. Stiffness is more than just duration and severity: a qualitative exploration in people with rheumatoid arthritis. Rheumatology (Oxford) 2015 Apr;54(4):615-622 [FREE Full text] [doi: 10.1093/rheumatology/keu379] [Medline: 25231178]
20. Hammarlund CS, Lexell J, Brogårdh C. Perceived consequences of ageing with late effects of polio and strategies for managing daily life: a qualitative study. BMC Geriatr 2017 Aug 9;17(1):179 [FREE Full text] [doi: 10.1186/s12877-017-0563-8] [Medline: 28793865]
21. Thille P, Ward N, Russell G. Self-management support in primary care: enactments, disruptions, and conversational consequences. Soc Sci Med 2014 May;108:97-105. [doi: 10.1016/j.socscimed.2014.02.041] [Medline: 24632054]
22. Crowley MJ, Holleman R, Klamerus ML, Bosworth HB, Edelman D, Heisler M. Factors associated with persistent poorly controlled diabetes mellitus: clues to improving management in patients with resistant poor control. Chronic Illn 2014 Dec;10(4):291-302 [FREE Full text] [doi: 10.1177/1742395314523653] [Medline: 24567193]
23. Matarese M, Lommi M, de Marinis MG, Riegel B. A systematic review and integration of concept analyses of self-care and related concepts. J Nurs Scholarsh 2018 May;50(3):296-305. [doi: 10.1111/jnu.12385] [Medline: 29645402]
24. O'Hara L, de Souza LH, Ide L. A Delphi study of self-care in a community population of people with multiple sclerosis. Clin Rehabil 2000 Feb;14(1):62-71. [doi: 10.1191/026921500666135189] [Medline: 10688346]
25. Learmonth YC, Motl RW. Physical activity and exercise training in multiple sclerosis: a review and content analysis of qualitative research identifying perceived determinants and consequences. Disabil Rehabil 2016;38(13):1227-1242. [doi: 10.3109/09638288.2015.1077397] [Medline: 26314587]
26. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
27. Yin K, Laranjo L, Tong HL, Lau AY, Kocaballi AB, Martin P, et al. Context-aware systems for chronic disease patients: scoping review. J Med Internet Res 2019 Jun 17;21(6):e10896 [FREE Full text] [doi: 10.2196/10896] [Medline: 31210138]
28. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 2;20(1):37-46 [FREE Full text] [doi: 10.1177/001316446002000104]
29. Corbin J, Strauss A. Managing chronic illness at home: three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/bf00989485]
30. Corbin JM, Strauss A. Unending Work and Care: Managing Chronic Illness at Home. San Francisco, CA: Jossey-Bass; 1988.
31. Dack C, Ross J, Stevenson F, Pal K, Gubert E, Michie S, et al. A digital self-management intervention for adults with type 2 diabetes: combining theory, data and participatory design to develop HeLP-Diabetes. Internet Interv 2019 Sep;17:100241 [FREE Full text] [doi: 10.1016/j.invent.2019.100241] [Medline: 31372349]

32. Mickelson RS, Holden RJ. Capturing the Medication Management Work System of Older Adults Using a Digital Diary Method. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2017 Presented at: HFES'17; October 9-13, 2017; Austin, Texas p. 555-559. [doi: 10.1177/1541931213601622]
33. Schultz TJ, Kitson AL. Measuring the context of care in an Australian acute care hospital: a nurse survey. Implement Sci 2010 Aug 2;5:60 [FREE Full text] [doi: 10.1186/1748-5908-5-60] [Medline: 20673373]
34. Krawczyk M, Sawatzky R, Schick-Makaroff K, Stajduhar K, Öhlen J, Reimer-Kirkham S, et al. Micro-meso-macro practice tensions in using patient-reported outcome and experience measures in hospital palliative care. Qual Health Res 2018 Mar 1;29(4):510-521. [doi: 10.1177/1049732318761366] [Medline: 29542400]
35. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010 Nov 2;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
36. Sutherland HJ, Till JE. Quality of life assessments and levels of decision making: differentiating objectives. Qual Life Res 1993 Aug;2(4):297-303. [doi: 10.1007/bf00434801] [Medline: 8220364]
37. Smith T, McNeil K, Mitchell R, Boyle B, Ries N. A study of macro-, meso- and micro-barriers and enablers affecting extended scopes of practice: the case of rural nurse practitioners in Australia. BMC Nurs 2019;18:14 [FREE Full text] [doi: 10.1186/s12912-019-0337-z] [Medline: 30976197]
38. Caldwell S, Mays N. Studying policy implementation using a macro, meso and micro frame analysis: the case of the collaboration for leadership in applied health research & care (CLAHRC) programme nationally and in North West London. Health Res Policy Syst 2012 Oct 15;10:32 [FREE Full text] [doi: 10.1186/1478-4505-10-32] [Medline: 23067208]
39. Kapiriri L, Norheim O, Martin D. Priority setting at the micro-, meso- and macro-levels in Canada, Norway and Uganda. Health Policy 2007 Jun;82(1):78-94. [doi: 10.1016/j.healthpol.2006.09.001] [Medline: 17034898]
40. Close C, Sinclair M, Liddle D, Mc Cullough J, Hughes C. Women's experience of low back and/or pelvic pain (LBPP) during pregnancy. Midwifery 2016 Jun;37:1-8. [doi: 10.1016/j.midw.2016.03.013] [Medline: 27217231]
41. Chen K, Liu C, Shyu YL, Yeh S. Living with chronic obstructive pulmonary disease: the process of self-managing chronic obstructive pulmonary disease. J Nurs Res 2016 Sep;24(3):262-271. [doi: 10.1097/jnr.0000000000000152] [Medline: 27015593]
42. Thompson M. Occupations, habits, and routines: perspectives from persons with diabetes. Scand J Occup Ther 2014 Mar;21(2):153-160. [doi: 10.3109/11038128.2013.851278] [Medline: 24215481]
43. Boehmer KR, Gionfriddo MR, Rodriguez-Gutierrez R, Dabrh AM, Leppin AL, Hargraves I, et al. Patient capacity and constraints in the experience of chronic disease: a qualitative systematic review and thematic synthesis. BMC Fam Pract 2016 Sep 1;17:127 [FREE Full text] [doi: 10.1186/s12875-016-0525-9] [Medline: 27585439]
44. Bowling CB, Vandenberg AE, Phillips LS, McClellan WM, Johnson TM, Echt KV. Older patients' perspectives on managing complexity in CKD self-management. Clin J Am Soc Nephrol 2017 Apr 3;12(4):635-643 [FREE Full text] [doi: 10.2215/CJN.06850616] [Medline: 28389529]
45. Browne C, Salmon N, Kehoe M. Bladder dysfunction and quality of life for people with multiple sclerosis. Disabil Rehabil 2015;37(25):2350-2358. [doi: 10.3109/09638288.2015.1027007] [Medline: 25801920]
46. Bukhave EB, Huniche L. Activity problems in everyday life-patients' perspectives of hand osteoarthritis: 'try imagining what it would be like having no hands'. Disabil Rehabil 2014;36(19):1636-1643. [doi: 10.3109/09638288.2013.863390] [Medline: 24308906]
47. Cobussen-Boekhorst H, Hermeling E, Heesakkers J, van Gaal B. Patients' experience with intermittent catheterisation in everyday life. J Clin Nurs 2016 May;25(9-10):1253-1261. [doi: 10.1111/jocn.13146] [Medline: 26991436]
48. Czuber-Dochan W, Dibley LB, Terry H, Ream E, Norton C. The experience of fatigue in people with inflammatory bowel disease: an exploratory study. J Adv Nurs 2013 Sep;69(9):1987-1999. [doi: 10.1111/jan.12060] [Medline: 23215959]
49. Dahlviken RM, Fridlund B, Mathisen L. Women's experiences of Takotsubo cardiomyopathy in a short-term perspective-a qualitative content analysis. Scand J Caring Sci 2015 Jun;29(2):258-267. [doi: 10.1111/scs.12158] [Medline: 24953349]
50. Dehghanzadeh S, Nayeri ND, Varaei S, Kheirkhah J. Living with cardiac resynchronization therapy: challenges for people with heart failure. Nurs Health Sci 2017 Mar;19(1):112-118. [doi: 10.1111/nhs.12325] [Medline: 28097793]
51. Franklin ZC, Smith NC, Fowler NE. A qualitative investigation of factors that matter to individuals in the pain management process. Disabil Rehabil 2016 Sep;38(19):1934-1942. [doi: 10.3109/09638288.2015.1107782] [Medline: 26728636]
52. Gallacher KI, May CR, Langhorne P, Mair FS. A conceptual model of treatment burden and patient capacity in stroke. BMC Fam Pract 2018 Jan 9;19(1):9 [FREE Full text] [doi: 10.1186/s12875-017-0691-4] [Medline: 29316892]
53. Hunt T, Madigan S, Williams M, Olds T. Use of time in people with chronic obstructive pulmonary disease-a systematic review. Int J Chron Obstruct Pulmon Dis 2014;9:1377-1388 [FREE Full text] [doi: 10.2147/COPD.S74298] [Medline: 25548519]
54. Levin-Zamir D, Badarne S, Najami M, Noy SG, Poraz I, Shapira M, et al. The use of focus groups as a basis for planning and implementing culturally appropriate health promotion among people with diabetes in the Arab community. Glob Health Promot 2016 Mar;23(1):5-14. [doi: 10.1177/1757975914548200] [Medline: 25802303]
55. Lidén E, Björk-Brämberg E, Svensson S. The meaning of learning to live with medically unexplained symptoms as narrated by patients in primary care: a phenomenological-hermeneutic study. Int J Qual Stud Health Well-Being 2015;10:27191 [FREE Full text] [doi: 10.3402/qhw.v10.27191] [Medline: 25887965]

56. Lindquist H, Enblom A, Bergmark K. Self-care among female cancer survivors with secondary lymphedema: a qualitative study. Rehabil Oncol 2015;33(1):51-57. [doi: 10.1097/01893697-201533010-00009]
57. Martin F, Kiwanuka T, Kawuma R, Zalwango F, Seeley J. Tasks and strategies of self-management of living with antiretroviral therapy in Uganda. AIDS Patient Care STDS 2013 Dec;27(12):697-706. [doi: 10.1089/apc.2013.0254] [Medline: 24219780]
58. Au T, Wong M, McMillan A, Bridges S, McGrath C. Treatment seeking behaviour in southern Chinese elders with chronic orofacial pain: a qualitative study. BMC Oral Health 2014 Jan 25;14:8 [FREE Full text] [doi: 10.1186/1472-6831-14-8] [Medline: 24460663]
59. Burnett K, Glaspole I, Holland AE. Understanding the patient's experience of care in idiopathic pulmonary fibrosis. Respirology 2019 Mar;24(3):270-277 [FREE Full text] [doi: 10.1111/resp.13414] [Medline: 30282111]
60. Cameron MH, Asano M, Bourdette D, Finlayson ML. People with multiple sclerosis use many fall prevention strategies but still fall frequently. Arch Phys Med Rehabil 2013 Aug;94(8):1562-1566. [doi: 10.1016/j.apmr.2013.01.021] [Medline: 23391522]
61. Cheung MM, Saini B, Smith L. Drawing asthma: an exploration of patients' perceptions and experiences. J Asthma 2018 Mar;55(3):284-293. [doi: 10.1080/02770903.2017.1325492] [Medline: 28514199]
62. Dial M, Holmes J, McGownd R, Wendler M. 'I do the best I can:' personal care preferences of patients of size. Appl Nurs Res 2018 Feb;39:259-264. [doi: 10.1016/j.apnr.2017.11.028] [Medline: 29422169]
63. Graham LJ, Connelly DM. 'Any movement at all is exercise': a focused ethnography of rural community-dwelling older adults' perceptions and experiences of exercise as self-care. Physiother Can 2013;65(4):333-341 [FREE Full text] [doi: 10.3138/ptc.2012-31] [Medline: 24396160]
64. Jacobsson LR, Milberg A, Hjelm K, Friedrichsen M. Experiences and own management regarding residual symptoms among people with Coeliac disease. Appl Nurs Res 2017 Jun;35:53-58. [doi: 10.1016/j.apnr.2017.02.009] [Medline: 28532727]
65. Jani B, Blane D, Browne S, Montori V, May C, Shippee N, et al. Identifying treatment burden as an important concept for end of life care in those with advanced heart failure. Curr Opin Support Palliat Care 2013 Mar;7(1):3-7. [doi: 10.1097/SPC.0b013e32835c071f] [Medline: 23196381]
66. Kawi J. Chronic low back pain patients' perceptions on self-management, self-management support, and functional ability. Pain Manag Nurs 2014 Mar;15(1):258-264. [doi: 10.1016/j.pmn.2012.09.003] [Medline: 23232149]
67. Moore L, Frost J, Britten N. Context and complexity: the meaning of self-management for older adults with heart disease. Sociol Health Illn 2015 Nov;37(8):1254-1269. [doi: 10.1111/1467-9566.12316] [Medline: 26235674]
68. Mousavizadeh S, Ashktorab T, Ahmadi F, Zandi M. From negligence to perception of complexities in adherence to treatment process in people with diabetes: a grounded theory study. Iran J Med Sci 2018 Mar;43(2):150-157 [FREE Full text] [Medline: 29749983]
69. O'Conor R, Martynenko M, Gagnon M, Hauser D, Young E, Lurio J, Supporting Asthma Self-Management Behaviors Among Aging Adults (SAMBA) Investigators. A qualitative investigation of the impact of asthma and self-management strategies among older adults. J Asthma 2017 Jan 2;54(1):39-45. [doi: 10.1080/02770903.2016.1193602] [Medline: 27315570]
70. Pauling JD, Domsic RT, Saketkoo LA, Almeida C, Withey J, Jay H, et al. Multinational qualitative research study exploring the patient experience of Raynaud's phenomenon in systemic sclerosis. Arthritis Care Res (Hoboken) 2018 Sep;70(9):1373-1384 [FREE Full text] [doi: 10.1002/acr.23475] [Medline: 29473715]
71. Rintala T, Paavilainen E, Åstedt-Kurki P. Everyday life of a family with diabetes as described by adults with type 1 diabetes. Eur Diabetes Nurs 2015 Feb 17;10(3):86-90 [FREE Full text] [doi: 10.1002/edn.234]
72. Schjoedt I, Sommer I, Bjerrum MB. Experiences and management of fatigue in everyday life among adult patients living with heart failure: a systematic review of qualitative evidence. JBI Database System Rev Implement Rep 2016 Mar;14(3):68-115. [doi: 10.11124/JBISRIR-2016-2441] [Medline: 27532141]
73. Stridsman C, Lindberg A, Skär L. Fatigue in chronic obstructive pulmonary disease: a qualitative study of people's experiences. Scand J Caring Sci 2014 Mar;28(1):130-138. [doi: 10.1111/scs.12033] [Medline: 23517049]
74. Walthall H, Jenkinson C, Boulton M. Living with breathlessness in chronic heart failure: a qualitative study. J Clin Nurs 2017 Jul;26(13-14):2036-2044. [doi: 10.1111/jocn.13615] [Medline: 27731919]
75. Zhang L, Gallagher R, Lowres N, Orchard J, Freedman SB, Neubeck L. Using the 'think aloud' technique to explore quality of life issues during standard quality-of-life questionnaires in patients with atrial fibrillation. Heart Lung Circ 2017 Feb;26(2):150-156. [doi: 10.1016/j.hlc.2016.05.121] [Medline: 27469897]
76. Eilertsen G, Ormstad H, Kirkevold M. Experiences of poststroke fatigue: qualitative meta-synthesis. J Adv Nurs 2013 Mar;69(3):514-525. [doi: 10.1111/jan.12002] [Medline: 22943761]
77. Kjeken I, Darre S, Slatkowsky-Cristensen B, Hermann M, Nilsen T, Eriksen CS, et al. Self-management strategies to support performance of daily activities in hand osteoarthritis. Scand J Occup Ther 2013 Jan;20(1):29-36. [doi: 10.3109/11038128.2012.661457] [Medline: 22376127]
78. Matthie N, Hamilton J, Wells D, Jenerette C. Perceptions of young adults with sickle cell disease concerning their disease experience. J Adv Nurs 2016 Jun;72(6):1441-1451 [FREE Full text] [doi: 10.1111/jan.12760] [Medline: 26350494]

79. dal Bello-Haas VP, O'Connell ME, Morgan D. Maintaining health and wellness in the face of dementia: an exploratory analysis of individuals attending a rural and remote memory clinic. Rural Remote Health 2014;14(3):2722 [FREE Full text] [Medline: 25100163]
80. Johnston C, Oprescu F, Gray M. Building the evidence for CRPS research from a lived experience perspective. Scand J Pain 2015 Oct 1;9(1):30-37. [doi: 10.1016/j.sjpain.2015.06.003] [Medline: 29911652]
81. Kneck A, Fagerberg I, Eriksson LE, Lundman B. Living with diabetes-development of learning patterns over a 3-year period. Int J Qual Stud Health Well-Being 2014;9:24375 [FREE Full text] [doi: 10.3402/qhw.v9.24375] [Medline: 25030359]
82. Matima R, Murphy K, Levitt N, BeLue R, Oni T. A qualitative study on the experiences and perspectives of public sector patients in Cape Town in managing the workload of demands of HIV and type 2 diabetes multimorbidity. PLoS One 2018;13(3):e0194191 [FREE Full text] [doi: 10.1371/journal.pone.0194191] [Medline: 29538415]
83. Arman M, Hök J. Self-care follows from compassionate care-chronic pain patients' experience of integrative rehabilitation. Scand J Caring Sci 2016 Jun;30(2):374-381. [doi: 10.1111/scs.12258] [Medline: 26395196]
84. Coventry PA, Small N, Panagioti M, Adeyemi I, Bee P. Living with complexity; marshalling resources: a systematic review and qualitative meta-synthesis of lived experience of mental and physical multimorbidity. BMC Fam Pract 2015 Nov 24;16:171 [FREE Full text] [doi: 10.1186/s12875-015-0345-3] [Medline: 26597934]
85. Fritz H. The influence of daily routines on engaging in diabetes self-management. Scand J Occup Ther 2014 May;21(3):232-240 [FREE Full text] [doi: 10.3109/11038128.2013.868033] [Medline: 24329181]
86. Gerrish K, Naisby A, Ismail M. Experiences of the diagnosis and management of tuberculosis: a focused ethnography of Somali patients and healthcare professionals in the UK. J Adv Nurs 2013 Oct;69(10):2285-2294. [doi: 10.1111/jan.12112] [Medline: 23421790]
87. Graffigna G, Barello S, Libreri C, Bosio C. How to engage type-2 diabetic patients in their own health management: implications for clinical practice. BMC Public Health 2014 Jun 25;14:648 [FREE Full text] [doi: 10.1186/1471-2458-14-648] [Medline: 24966036]
88. Kimani K, Murray S, Grant L. Multidimensional needs of patients living and dying with heart failure in Kenya: a serial interview study. BMC Palliat Care 2018 Feb 17;17(1):28 [FREE Full text] [doi: 10.1186/s12904-018-0284-6] [Medline: 29454383]
89. Robinson K, Lucas E, van den Dolder P, Halcomb E. Living with chronic obstructive pulmonary disease: the stories of frequent attenders to the emergency department. J Clin Nurs 2018 Jan;27(1-2):48-56. [doi: 10.1111/jocn.13842] [Medline: 28382725]
90. Roberts AR, Adams KB, Warner CB. Effects of chronic illness on daily life and barriers to self-care for older women: a mixed-methods exploration. J Women Aging 2017;29(2):126-136. [doi: 10.1080/08952841.2015.1080539] [Medline: 27455030]
91. McQuoid J, Jowsey T, Talaulikar G. Contextualising renal patient routines: everyday space-time contexts, health service access, and wellbeing. Soc Sci Med 2017 Jun;183:142-150 [FREE Full text] [doi: 10.1016/j.socscimed.2017.04.043] [Medline: 28482275]
92. Apps L, Hewitt S, Green R, Bradding P, Murphy A, Martin N, et al. P245 '… no cleaning, no stairs, no sex…everything just stops': understanding living with severe asthma to inform effective self-management. Thorax 2015 Nov 12;70(Suppl 3):A200.1. [doi: 10.1136/thoraxjnl-2015-207770.381]
93. Ferreira S, Cordeiro R, Cajuhy F. Vulnerability in adults with sickle cell disease: subsidies for nursing care. Ciencia 2013;12(4):711-718. [doi: 10.4025/cienccuidsaude.v12i4.18723]
94. Lindgren I, Gard G, Brogårdh C. Shoulder pain after stroke-experiences, consequences in daily life and effects of interventions: a qualitative study. Disabil Rehabil 2018 May;40(10):1176-1182. [doi: 10.1080/09638288.2017.1290699] [Medline: 28637154]
95. Westra BL, Paitich N, Ekstrom D, Mehle SC, Kaeding M, Abdo S, et al. Getting on with living life: experiences of older adults after home care. Home Healthc Nurse 2013 Oct;31(9):493-501; quiz 501 [FREE Full text] [doi: 10.1097/NHH.0b013e3182a87654] [Medline: 24081131]
96. Bratzke LC, Muehrer RJ, Kehl KA, Lee KS, Ward EC, Kwekkeboom KL. Self-management priority setting and decision-making in adults with multimorbidity: a narrative review of literature. Int J Nurs Stud 2015 Mar;52(3):744-755 [FREE Full text] [doi: 10.1016/j.ijnurstu.2014.10.010] [Medline: 25468131]
97. Wilson O, Kirwan J, Dures E, Quest E, Hewlett S. The experience of foot problems and decisions to access foot care in patients with rheumatoid arthritis: a qualitative study. J Foot Ankle Res 2017;10(1):4 [FREE Full text] [doi: 10.1186/s13047-017-0188-3] [Medline: 28138340]
98. Peoples H, Brandt A, Wæhrens EE, la Cour K. Managing occupations in everyday life for people with advanced cancer living at home. Scand J Occup Ther 2017 Jan;24(1):57-64. [doi: 10.1080/11038128.2016.1225815] [Medline: 27578556]
99. Swenne CL, Jangland E, Arakelian E. Patients' experiences of their everyday life 14 months after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy-a qualitative follow-up study. Scand J Caring Sci 2017 Dec;31(4):904-913. [doi: 10.1111/scs.12412] [Medline: 28124449]
100. Karsh B, Waterson P, Holden RJ. Crossing levels in systems ergonomics: a framework to support 'mesoergonomic' inquiry. Appl Ergon 2014 Jan;45(1):45-54. [doi: 10.1016/j.apergo.2013.04.021] [Medline: 23706573]

101. Mickelson RS, Unertl KM, Holden RJ. Medication management: the macrocognitive workflow of older adults with heart failure. JMIR Hum Factors 2016 Oct 12;3(2):e27 [FREE Full text] [doi: 10.2196/humanfactors.6338] [Medline: 27733331]
102. Lambert K, Mansfield K, Mullan J. How do patients and carers make sense of renal dietary advice? A qualitative exploration. J Ren Care 2018 Dec;44(4):238-250. [doi: 10.1111/jorc.12260] [Medline: 30259677]
103. Mamykina L, Smaldone A, Bakken SR. Adopting the sensemaking perspective for chronic disease self-management. J Biomed Inform 2015 Aug;56:406-417 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.006] [Medline: 26071681]
104. Gorman RK, Wellbeloved-Stone CA, Valdez RS. Uncovering the invisible patient work system through a case study of breast cancer self-management. Ergonomics 2018 Dec;61(12):1575-1590. [doi: 10.1080/00140139.2018.1503339] [Medline: 30044709]
105. Breitbart W, Pessin H, Rosenfeld B, Applebaum AJ, Lichtenthal WG, Li Y, et al. Individual meaning-centered psychotherapy for the treatment of psychological and existential distress: a randomized controlled trial in patients with advanced cancer. Cancer 2018 Aug 1;124(15):3231-3239 [FREE Full text] [doi: 10.1002/cncr.31539] [Medline: 29757459]
106. McCullough A, Ruehrdanz A, Jenkins M, Gilmer M, Olson J, Pawar A, et al. Measuring the effects of an animal-assisted intervention for pediatric oncology patients and their parents: a multisite randomized controlled trial [formula: see text]. J Pediatr Oncol Nurs 2018 May;35(3):159-177. [doi: 10.1177/1043454217748586] [Medline: 29268667]
107. van Houtum L, Rijken M, Heijmans M, Groenewegen P. Patient-perceived self-management tasks and support needs of people with chronic illness: generic or disease specific? Ann Behav Med 2015 Apr;49(2):221-229. [doi: 10.1007/s12160-014-9649-0] [Medline: 25199663]
108. Kuo C, Liang S, Tsay S, Wang T, Cheng S. Symptom management tasks and behaviors related to chemotherapy in Taiwanese outpatients with breast cancer. Eur J Oncol Nurs 2015 Dec;19(6):654-659. [doi: 10.1016/j.ejon.2015.03.012] [Medline: 25959612]
109. Liang S, Yates P, Edwards H, Tsay S. Opioid-taking tasks and behaviours in Taiwanese outpatients with cancer. J Clin Nurs 2008 Aug;17(15):2079-2088. [doi: 10.1111/j.1365-2702.2008.02332.x] [Medline: 18705784]
110. Pozuelo-Carrascosa DP, Martínez-Vizcaíno V, Sánchez-López M, Bartolomé-Gutiérrez R, Rodríguez-Martín B, Notario-Pacheco B. Resilience as a mediator between cardiorespiratory fitness and mental health-related quality of life: a cross-sectional study. Nurs Health Sci 2017 Sep;19(3):316-321. [doi: 10.1111/nhs.12347] [Medline: 28590081]
111. Christensen AJ. Patient-by-treatment context interaction in chronic disease: a conceptual framework for the study of patient adherence. Psychosom Med 2000;62(3):435-443. [doi: 10.1097/00006842-200005000-00017] [Medline: 10845357]
112. Steca P, D'Addario M, Magrin M, Miglioretti M, Monzani D, Pancani L, et al. A type a and type d combined personality typology in essential hypertension and acute coronary syndrome patients: associations with demographic, psychological, clinical, and lifestyle indicators. PLoS One 2016;11(9):e0161840 [FREE Full text] [doi: 10.1371/journal.pone.0161840] [Medline: 27589065]
113. Moryś JM, Bellwon J, Jeżewska M, Adamczyk K, Gruchała M. The evaluation of stress coping styles and type D personality in patients with coronary artery disease. Kardiol Pol 2015;73(7):557-566 [FREE Full text] [doi: 10.5603/KP.a2015.0039] [Medline: 25733180]

Edited by G Eysenbach; submitted 11.10.19; peer-reviewed by R Mickelson, H Mehdizadeh, R Cruz Martínez; comments to author 12.12.19; revised version received 04.02.20; accepted 29.03.20; published 02.06.20.

Please cite as:
Yin K, Jung J, Coiera E, Laranjo L, Blandford A, Khoja A, Tai WT, Phillips DP, Lau AYS
Patient Work and Their Contexts: Scoping Review
J Med Internet Res 2020;22(6):e16656
URL: https://www.jmir.org/2020/6/e16656
doi:10.2196/16656
PMID:32484449

Original Paper

Physiotherapy With Telerehabilitation in Patients With Complicated Postoperative Recovery After Esophageal Cancer Surgery: Feasibility Study

Maarten A van Egmond[1,2,3*], PT, MSc; Raoul H H Engelbert[1,2*], PT, PhD; Jean H G Klinkenbijl[4,5*], MD, PhD; Mark I van Berge Henegouwen[6*], MD, PhD; Marike van der Schaaf[1,2*], PT, PhD

[1]Department of Rehabiltation, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands
[2]Faculty of Health, Center of Expertise Urban Vitality, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[3]Faculty of Health, European School of Physiotherapy, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[4]Department of Surgery, Gelre Hospital Apeldoorn-Zutphen, Apeldoorn, Netherlands
[5]University of Amsterdam, Amsterdam, Netherlands
[6]Department of Surgery, Cancer Center Amsterdam, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands
[*]all authors contributed equally

Corresponding Author:
Maarten A van Egmond, PT, MSc
Department of Rehabiltation
Amsterdam University Medical Centers
University of Amsterdam
Meibergdreef 9
Amsterdam, 1105AZ
Netherlands
Phone: 31 621157870
Email: m.a.vanegmond@amsterdamumc.nl

Abstract

Background: Improvement of functional status with physiotherapy is an important goal for patients with postoperative complications and an increased length of hospital stay (LoS) after esophagectomy. Supervised physiotherapy with telerehabilitation instead of conventional face-to-face care could be an alternative to treat these patients in their home environment after hospital discharge (T0), but its feasibility has not yet been investigated in detail.

Objective: The aim of this study was to investigate the feasibility of a 12-week supervised postoperative physiotherapy intervention with telerehabilitation for patients with esophageal cancer who underwent esophagectomy and had postoperative complications or who had an increased LoS. The secondary objective was to investigate the preliminary effectiveness of telerehabilitation on functional recovery compared with usual care.

Methods: A prospective feasibility study with a matched historical comparison group was performed. Feasibility outcomes included willingness and adherence to participate, refusal rate, treatment duration, occurrence of adverse events, and patient satisfaction. Secondary outcome measures were measurements of musculoskeletal and cardiovascular functions and activities according to the domains of the International Classification of Functioning, Disability and Health.

Results: A total of 22 patients with esophageal cancer who underwent esophagectomy and had postoperative complications or an increased LoS were included. The mean age at surgery was 64.55 (SD 6.72) years, and 77% (17/22) of patients were male. Moreover, 15 patients completed the intervention. Patient adherence was 99.8% in the first 6 weeks and dropped to 75.6% in the following 6 weeks, with a mean difference of −24.3% (95% CI 1.3 to 47.2; *P*=.04). At 3 months post operation, no differences in functional status were found between the intervention group and the matched historical comparison group.

Conclusions: This study showed that a postoperative physiotherapeutic intervention with telerehabilitation is feasible for patients with postoperative complications or an increased LoS after esophageal cancer surgery up to 6 weeks after T0.

(J Med Internet Res 2020;22(6):e16056) doi:10.2196/16056

KEYWORDS

physical therapy modalities; telerehabilitation; telemedicine; esophageal neoplasms; surgical oncology

Introduction

Background

Surgical resection of the esophagus is the primary curative treatment for patients with esophageal cancer and is associated with a high risk of postoperative complications, varying from 25% to 60% [1,2]. This leads to an increased length of hospital stay (LoS) and a delayed postoperative recovery, with a significant decline in physical function in the first 3 months after surgery [2,3].

It has been demonstrated in many surgical populations that improving preoperative functional status by exercise training had a positive effect on long-term postoperative outcomes [4,5]. However, recent studies have shown that preoperative functional status was not associated with postoperative complications in patients treated with esophagectomy, justifying the need to focus on treating these patients in the postoperative phase [6,7].

Patients with postoperative complications after esophagectomy often have fatigue, decreased exercise capacity, and disability such as impaired walking capacity and their recovery could take up to one year and beyond [3,8]. These symptoms are explained by altered cardiopulmonary function, generalized muscle weakness, and malnutrition, and physiotherapists play an important role in improving these aspects of physical functioning [9].

Telerehabilitation as an Alternative to Face-to-Face Care

Instead of face-to-face care, postoperative physiotherapy can also be streamed by telerehabilitation. Telerehabilitation is a medium to provide physiotherapy with electronic health (eHealth), defined as *the delivery of rehabilitation services to patients at a distance using information and communication technologies* [10]. Telerehabilitation has shown to be a valuable tool in improving postoperative outcomes and functional recovery in surgical patients, where patients considered reduced barriers for travel, flexible exercise hours, and the ability to directly integrate exercises into daily life as positive [11,12].

Moreover, telerehabilitation interventions have been valuable to overcome discontinuities that may arise in communication between hospital and primary care, where physiotherapists may have a lack of knowledge about how to treat patients after a highly complex surgery [13].

There is evidence showing positive effects of physiotherapy with telerehabilitation on clinical outcomes in patients with cancer, patients with cardiac disease, and patients with musculoskeletal disorders, but information on the feasibility of this intervention in the postoperative phase of patients with esophageal cancer treated with esophagectomy is lacking [14,15].

Objectives

Therefore, the primary objective of this prospective feasibility study was to investigate the feasibility of a 12-week supervised postoperative telerehabilitation program for patients with esophageal cancer who underwent esophagectomy and had postoperative complications or who had an increased LoS. The secondary objective was to investigate the preliminary effectiveness of telerehabilitation on functional recovery compared with a matched historical comparison group receiving usual care.

Methods

Ethical Approval

The Medical Ethical Committee (METC) of the Amsterdam University Medical Centers provided ethical approval for this study (NL58388.018.16). All patients provided written informed consent. As this was a feasibility study, sample size calculations have not been performed, and the initial sample size of 30 participants was pragmatically chosen. Patients could leave the study at any time for any reason if they wished to do so without any consequences.

Study Design

A prospective feasibility study was performed in patients treated with esophagectomy. To assess preliminary effectiveness, the patients who underwent the complete treatment were matched with a historical comparison group of patients who underwent esophagectomy and had postoperative complications, receiving usual face-to-face care between March 2012 and October 2014. We decided to match one case to 2 patients from a historical comparison group to optimize statistical power. Data collected from this historical comparison group were part of a previous study performed by the same research group, from which the METC waived the need for informed consent [6]. Patients were matched for gender, age, American Society of Anesthesiologists Physical Status Classification, comorbidities, Body Mass Index, pulmonary function, surgical procedure, and severity of postoperative complications.

Participants

Patients were recruited from the surgical wards at the Gastrointestinal Oncologic Centre Amsterdam of the Amsterdam University Medical Centre, located in the Academic Medical Centre, just before discharge from the hospital by the supervising physiotherapist or the investigator. Patients who refused to participate were referred to face-to-face physiotherapy in primary care.

Inclusion Criteria

Participants were included if they were aged 18 years or older and the primary reason of hospital stay was status after esophagectomy, they had internet access at home, and they signed the informed consent form. Moreover, participants were included if they had a postoperative complication, grade 3a to 4 according to the Clavien-Dindo classification. This 5-scale classification reports surgical complications based on the type of therapy required to treat the complication [16]. Participants were also included if the postoperative LoS was longer than 9 days because they were physically too weak to be discharged earlier. There was an indication for face-to-face physiotherapy in primary care if a patient was not yet able to walk or transfer independently because of loss of muscle strength, mobility, or balance at discharge.

Exclusion Criteria

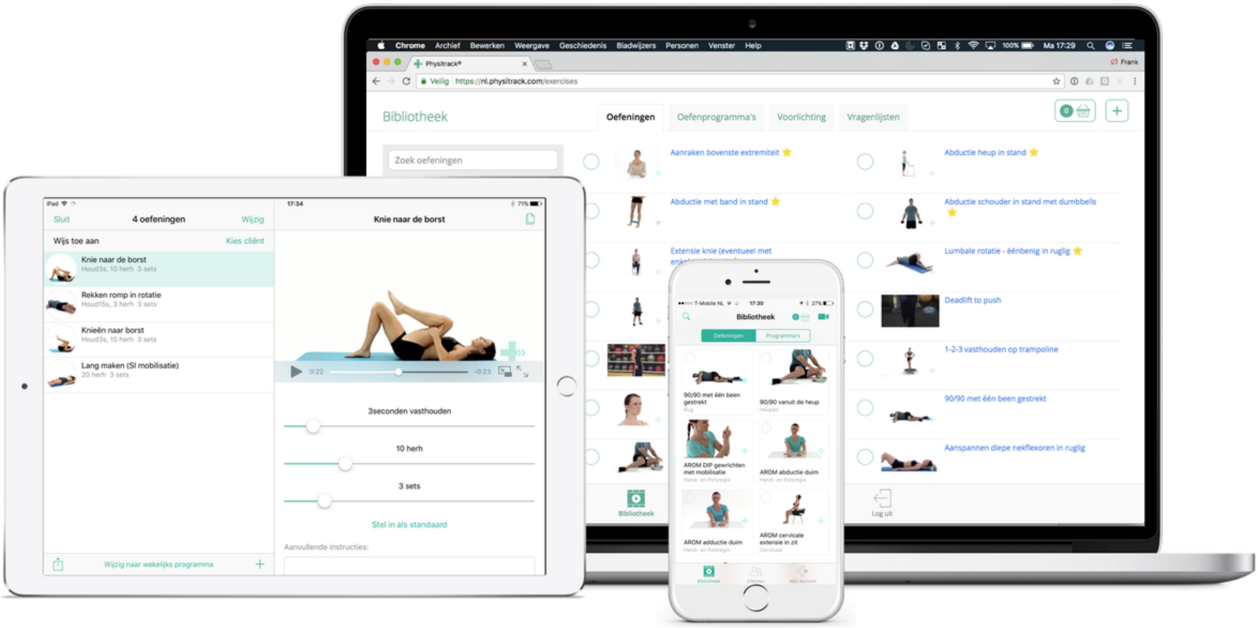
Patients were excluded if they were unable to complete self-reported questionnaires, insufficiently able to read or speak Dutch, had cognitive disorders, or had any other severe medical conditions that prevented them from doing unsupervised exercises at home.

Intervention

Participants received a 12-week supervised home-based telerehabilitation intervention after hospital discharge (T0) in their home environment. Before T0, a physiotherapist from the surgical ward instructed the patient on the telerehabilitation intervention.

The telerehabilitation intervention was provided with Physitrack (Physitrack Limited). Physitrack is an eHealth platform that enables physiotherapists to design home exercise programs and track patient adherence. Patients were provided with a goal-oriented exercise program created by the physiotherapist that could be accessed by a tablet, mobile phone, or computer (Figure 1). The physiotherapist accurately monitored the progress of the patient in weekly telephone, email, or video sessions, and exercises were adapted via the eHealth platform if needed. Physitrack had provided their services for this research project free of charge, and they will use the outcomes of this study to improve their services. They were not involved in the design, execution, analysis, and conclusions of this research. Physitrack will only have access to the published paper with its results, with no access to raw data.

Figure 1. Goal-oriented exercise program created by the physiotherapist, accessible by tablet, mobile phone, or computer (Used with permission from Physitrack).



The postoperative physiotherapeutic intervention with telerehabilitation was aimed at improving functional status. The intervention lasted 12 weeks with at least two sessions per week, depending on whether the treatment goals were achieved. The exercises were tailored to the patients' specific condition and needs, which were determined a day before T0. The physiotherapy goals were determined by using the patient-specific complaint list [17]. The exercises were aimed at improving the functional activity level of the patient, by increasing muscle strength, coordination, range of joint motion, and stamina. The intensity and frequency of the functional exercises were determined according to the guidelines of the American College of Sports Medicine [18]. Cardiorespiratory exercises to improve stamina were performed on a moderate-to-vigorous intensity level, measured using the Borg rating of perceived exertion scale (scores 6-20), for at least two sessions per week. Rating of perceived exertion with the Borg scale is a generally used and reliable scale to monitor and evaluate exercise intensity. A score from 13 to 16 relates to the moderate-to-vigorous intensity level, and this allowed us to monitor and adapt the appropriate intensity [19]. Exercises to improve muscle strength were performed 2 to 3 days per week on 60% to 70% of the 1 repetition maximum (moderate-to-hard intensity). We used the Holten curve that relates the percentage of the 1 repetition maximum to the estimated repetitions of that intensity. This allowed us to adapt the exercises without using fitness equipment to measure the 1 repetition maximum directly [18,20].

Feasibility Outcome Measures

Feasibility outcome measures were calculated for the 15 patients who completed the 12-week supervised home-based telerehabilitation intervention. Feasibility outcomes included refusal rate; adherence to the telerehabilitation intervention operationalized in the amount and duration of email, phone, and video calls conducted by patients and physiotherapists; treatment duration per session; adverse events; and patient satisfaction. Patient satisfaction was recorded with a modified telemedicine satisfaction and usefulness questionnaire (TSUQ), a 30-item Likert-type questionnaire including 3 subscales (usefulness, communication, and user friendliness) at 6 weeks post operation (T1) and at 3 months post operation (T2) [21]. Scores range from 30 to 150, with high scores indicating a higher satisfaction.

The telerehabilitation intervention was considered as feasible if at least an 80% adherence rate was achieved, if no adverse events took place, and if the average total patient satisfaction was higher than 75% (score >120).

Effectiveness Outcome Measures

Secondary outcome measures on preliminary effectiveness were musculoskeletal and cardiovascular functioning and level of activities according to the domains of the International Classification of Functioning, Disability and Health [22].

Handgrip strength was measured using the Jamar grip strength dynamometer (Lafayette Instrument Company) as a measure of generalized muscle strength [23,24]. Maximal inspiratory pressure was measured as an indicator of inspiratory muscle strength, with a Micro Respiratory Pressure Meter [2,4]. Functional lower extremity muscle function was measured with the 30-second chair stand test (30CST). This test measures extremity strength in relation to demanding functional daily activities such as stair climbing and getting out of a chair [25]. Walking capacity was measured using the 2-min walk test (2MWT) [26].

Fatigue was measured using the Multidimensional Fatigue Inventory [27].

Self-reported activities were measured using the Longitudinal Ageing Study Amsterdam Physical Activity Questionnaire (LAPAQ) in which patients reported the type, frequency, and duration of daily activities in the past 14 days. Health-related quality of life (HRQL) was measured using the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30, version 3.0 [28].

The effectiveness of outcome measures was recorded before the start of the intervention (T0) and at T1 and T2.

Standardized operating procedures of all measurements were used to guarantee uniformity and accuracy in operationalization. Trained and experienced physiotherapists performed the standardized measurements.

Statistical Analysis

Data were analyzed in the Statistical Package for the Social Sciences (version 25.0; IBM SPSS Statistics for Windows, IBM Corp). Statistical tests were analyzed two sided and considered significant with an alpha value ≤.05.

Baseline characteristics were summarized with descriptive statistics, where discrete variables were expressed as counts with percentages, ordinal variables as median and interquartile ranges (P25-75), and continuous variables as mean and standard deviation, and in case of a skewed distribution, they were expressed as median and interquartile range. Differences in outcomes before and after the intervention were determined by using a paired samples *t* test or a Wilcoxon matched-pairs signed rank test for skewed data. Differences between the intervention group and the historical comparison group were tested using a linear mixed model analysis to account for the dependency between observations.

Results

Baseline Characteristics

From January 2017 to October 2018, 22 patients with esophageal cancer who underwent esophagectomy were included in the study after obtaining informed consent. The study was terminated after the inclusion of the 22nd patient because we reached a point in data collection after which no new or relevant information emerged with respect to answering the primary research question.

The mean age at surgery was 64.6 (SD 6.7) years, and 77% (17/22) of patients were male. All patients received neoadjuvant chemoradiation therapy. At enrollment, mean pulmonary function expressed as a percent score of the predicted pulmonary function value was 116.1 (SD 18.7) for forced vital capacity, 109.0 (SD 19.3) for forced expiratory volume in 1 second, and 109.2 (SD 30.1) for inspiratory vital capacity. Except for 2 patients, all other patients were surgically treated with a minimally invasive transthoracic esophagectomy. In addition, 36% (8/22) of patients had a hospital stay of more than 9 days. Moreover, 91% (20/22) of patients had postoperative complications, of which 70% (14/22) required a surgical, an endoscopic, or a radiological intervention. Patient characteristics are presented in Table 1.

Table 1. Characteristics of the entire study population.

Patient characteristics	Study population (N=22)
Gender (male), n (%)	17 (77)
ASA[a] classification, n (%)	
I[b]	5 (22)
II[c]	10 (45)
III[d]	7 (31)
Age (years), mean (SD)	64.6 (6.7)
BMI[e], mean (SD)	26.5 (4.4)
Comorbidities, n (%)	
Cardiovascular	7 (31)
COPD[f]	0 (0)
DM II[g]	2 (9)
Cigarette smoking	2 (9)
Pulmonary function (percent predicted), mean (SD)	
FVC[h]	116.1 (18.7)
FEV_1[i]	109.0 (19.3)
IVC[j]	109.2 (30.1)
Surgical procedure, n (%)	
Transhiatal open	0 (0)
Transhiatal minimally invasive	1 (5)
Transthoracal open	0 (0)
Transthoracal minimally invasive	20 (91)
Esophageal resection with colon interposition	1 (5)
Clavien-Dindo postoperative complications, n (%)	
No complications	2 (9)
Grade 1	2 (9)
Grade 2	4 (18)
Grade 3a	3 (14)
Grade 3b	4 (18)
Grade 4a	7 (32)
Grade 4b	0 (0)
Grade 5	0 (0)

[a] ASA: American Society of Anesthesiologists.

[b] I: healthy person.

[c] II: mild systemic disease.

[d] III: severe systemic disease.

[e] BMI: body mass index is calculated as weight in kilograms divided by height in meters squared.

[f] COPD: chronic obstructive pulmonary disease.

[g] DM II: diabetes mellitus type 2.

[h] FVC: functional vital capacity.

[i] FEV1: forced expiratory volume in the first second of expiration.

[j] IVC: inspiratory vital capacity.

Feasibility

Of 22 patients, a total of 15 (68%) patients completed the 12-week program. Of the 7 patients who did not complete the study, 1 was discharged to a nursing home after inclusion, 2 quit the study intervention after 3 and 4 weeks because they preferred face-to-face physiotherapy, and 4 patients were withdrawn by the investigator because postoperative treatment required a multidisciplinary approach (n=3) or because of the presence of metastases (n=1). These patients did not systematically differ in baseline characteristics from the patients who completed the program.

The average duration of the treatment program was 11.1 (SD 5.2) weeks. Of the 4671 exercises provided to patients, 1337 (28.62%) were aimed at lower extremity muscle strength, 996 (21.32%) were aimed at respiration, and 1150 (24.62%) were aimed at walking.

Patient adherence, operationalized in the performance rate of exercises to the telerehabilitation intervention, was 99.8% in the first 6 weeks and dropped to 75.6% in the following 6 weeks, with a mean difference of −24.3% (95% CI 1.3 to 47.2; *P*=.04). The accomplishment of treatment goals was the main reason reported for being less or not adherent to the program anymore.

The physiotherapist and patients contacted each other 204 times in 243 weeks, with a minimum of 1 and a maximum of 3 times a week for coaching, for regular follow-ups, and for adjusting the treatment program, dependent on the patient's needs. Of these 204 direct patient contacts, 1 (0.5%) took place with a video connection, 26 (12.7%) with email, 122 (59.8%) with telephone, and 55 (27.0%) with live contact via home visits.

Total average patient satisfaction (range 30-150) measured at T1 was 135.0 (SD 19.5), with subscores on usefulness (range 10-50) being 44.66 (SD 7.4), communication (range 11-55) being 48.3 (SD 8.1), and user friendliness (range 9-45) being 42.8 (SD 3.2). Patients appreciated weekly follow-ups by telephone or email and especially appreciated the flexibility they had to perform the exercises at home. They rated the telerehabilitation app as user friendly, and they did not miss the physical presence of the physiotherapist to follow the exercise program. No adverse events took place during measurements or exercise sessions. Total average patient satisfaction at T2 was 139.6 (SD 15.4). Textbox 1 provides a selection of quotes provided by participants more than once about experiences with the program.

Textbox 1. Patient experiences.

Quotes:

- "It gave a lot of confidence to work at home on my recovery with supervision of a PT" [Mrs S, 70 years]
- "I could do the exercises whenever I wanted, that was very convenient" [Mr W, 54 years]
- "Without this program I would never have been that far" [Mr J, 66 years]
- "I should not have thought about going to the physiotherapist twice a week" [Mrs B, 60 years]
- "By practicing at home, I knew what I was doing it for. That was very motivating" [Mr B, 62 years]
- "I missed incentives in the program" [Mr B, 49 years]
- "I did not miss the physical presence of the physiotherapist, I felt that I could always reach him through the app" [Mrs B, 64 years]
- "Along the way, I found the exercise program less relevant, I could already do my daily activities again" [Mr S, 62 years]

Effectiveness

A total of 15 patients who completed the telerehabilitation program were matched with 30 patients from a historical comparison group for both pre- and postoperative characteristics (gender, age, preoperative pulmonary function, type of surgery, and postoperative complications classified according to Clavien-Dindo). Table 2 provides details about the matching characteristics.

At T0, patients in the intervention group had significantly lower functional capacity measures compared with reference values than patients in the matched historical comparison group (Table 3).

At 3 months post operation, no differences in functional status measures were found between the intervention group and the matched control group (Table 4).

Within the intervention group, 30CST, 2MWT, fatigue, and HRQL improved significantly between T0 and T1 and between T1 and T2, whereas activities of daily life (ADL) decreased significantly between T0 and T1 and improved again between T1 and T2 (Table 5).

Table 2. Patient characteristics of the intervention group matched with a historical comparison group.

Patient characteristics	Intervention (n=15)	Matched controls (n=30)
Gender (male), n (%)	11 (73)	22 (73)
ASA[a] classification, n (%)		
I[b]	3 (20)	5 (16)
II[c]	8 (53)	15 (50)
III[d]	4 (26)	10 (33)
Age (years), mean (SD)	62.8 (6.9)	60.3 (7)
BMI[e], mean (SD)	26.1 (3.5)	25.2 (4)
Comorbidities, n (%)		
Cardiovascular	6 (40)	5 (16)
COPD[f]	0 (0)	3 (10)
DM II[g]	1 (7)	1 (3)
Cigarette smoking	1 (7)	7 (23)
Pulmonary function (percent predicted), mean (SD)		
FVC[h]	115.0 (20.1)	116.3 (16.2)
FEV_1[i]	105.4 (20.1)	110.2 (20.7)
IVC[j]	114.1 (21.9)	112.0 (16.7)
Surgical procedure, n (%)		
Transhiatal open	0 (0)	0 (0)
Transhiatal minimally invasive	0 (0)	2 (7)
Transthoracal open	0 (0)	1 (3)
Transthoracal minimally invasive	14 (93)	27 (90)
Esophageal resection with colon interposition	1 (7)	0 (0)
Clavien-Dindo postoperative complications, n (%)		
No complications	2 (13)	11 (37)
Grade 1	2 (13)	4 (13)
Grade 2	2 (13)	7 (23)
Grade 3a	3 (20)	4 (13)
Grade 3b	2 (13)	1 (3)
Grade 4a	4 (27)	2 (7)
Grade 4b	0 (0)	1 (3)
Grade 5	0 (0)	0 (0)

[a] ASA: American Society of Anesthesiologists.

[b] I: healthy person.

[c] II: mild systemic disease.

[d] III: severe systemic disease.

[e] BMI: body mass index is calculated as weight in kilograms divided by height in meters squared.

[f] COPD: chronic obstructive pulmonary disease.

[g] DM II: diabetes mellitus type 2.

[h] FVC: functional vital capacity.

[i] FEV1: forced expiratory volume in the first second of expiration.

[j] IVC: inspiratory vital capacity.

Table 3. Functional status capacity outcome measures at hospital discharge (T0). Beta values represent the differences in functional status between the historical control group and the intervention group at T0.

Functional status outcome	Intervention	Control	Beta	95% CI	*P* value
RHGS[a] (percent predicted), mean (SD)	92.4 (19.7)	107.9 (23.2)	−15.5	−31.9 to 0.79	.04[b]
LHGS[c] (percent predicted), mean (SD)	97.1 (20.8)	106.2 (22.4)	−11.9	−26.6 to 2.9	.11[b]
30CST[d] (percent predicted), mean (SD)	50.8 (31.6)	89.0 (34.4)	−33.2	−53.8 to −12.7	.003[b]
2MWT[e] (meters), mean (SD)	117.4 (50.6)	154.4 (32.3)	−22.6	−42.7 to −2.5	.03[b]

[a] RHGS: right-hand grip strength.

[b] *P*≤.05 is considered significant.

[c] LHGS: left-hand grip strength.

[d] 30CST: 30-second chair stand test.

[e] 2MWT: 2-min walk test.

Table 4. Within-group differences between hospital discharge (T0) and 3 months post operation (T2) and between-group differences at T2 in measures of functional status. Within-group differences represent the differences in functional status between T0 and T2. Beta values represent the differences in functional status between the historical control group and the intervention group at T2.

Functional status outcome	Within-group differences (T0-T2)[a,b]				Between-group differences at T2	
	Intervention (n=15)		Historical control (n=30)		Beta	
	Mean (95% CI)	*P* value	Mean (95% CI)	*P* value	Mean (95% CI)	*P* value
LHGS[c]	10.4 (0.1 to 20.8)	.048[d]	−4.1 (−8.7 to 0.5)	.08	0.8 (14.2 to −12.7)	.91
RHGS[e]	12.3 (0.9 to 23.7)	.04[d]	−3.2 (−8.9 to 2.4)	.25	−1.0 (−15.3 to 13.3)	.89
MIP[f,g]	—[h]	—	—	—	13.7 (−14.0 to 41.4)	.32
30CST[i]	69.7 (51.6 to 87.8)	<.001[d]	29.8 (18.7 to 40.9)	<.001[d]	5.9 (−15.3 to 27.0)	.58
2MWT[j]	82.4 (53.4 to 111.3)	.001[d]	41.2 (27.3 to 55.1)	<.001[d]	16.8 (−7.6 to 41.2)	.17
ADL[g,k]	—	—	—	—	−444.3 (−1417.0 to 528.3)	.36
Fatigue[g]	—	—	—	—	−3.6 (−16.0 to 8.8)	.55
HRQL[l]	—	—	—	—	3.5 (−9.0 to 16.11)	.57

[a] T0: hospital discharge.

[b] T2: 3 months post operation.

[c] LHGS: left-hand grip strength.

[d] *P*<.05 is considered significant.

[e] RHGS: right-hand grip strength.

[f] MIP: maximal inspiratory pressure.

[g] These measurements were not performed at T0 and therefore were excluded from this analysis.

[h] Missing data.

[i] 30CST: 30-second chair stand test.

[j] 2MWT: 2-min walk test.

[k] ADL: activities of daily life.

[l] HRQL: health-related quality of life.

Table 5. Mean differences in functional status outcomes between hospital discharge and 6 weeks post operation (T1) and between T1 and 3 months post operation in the intervention group (n=15).

Measurements	Δ[a]T0-T1[b,c] (95% CI)	*P* value	ΔT1-T2[d] (95% CI)	*P* value
RHGS[e]	7.4 (−5.1 to 19.8)	.22	5.1 (−1.5 to 11.6)	.12
LHGS[f]	9.6 (−0.6 to 19.8)	.06	1.0 (−5.0 to 6.9)	.74
MIP[g,h]	—[i]	—	9.6 (−1.1 to 20.3)	.07
30CST[j]	53.0 (38.5 to 67.5)	<.001[k]	19.0 (10.2 to 27.9)	.001[k]
2 MWT (m)[l]	51.0 (21.9 to 80.2)	.002[k]	30.3 (15.5 to 445.0)	.001[k]
MFI[m] fatigue	−10.2 (−16.8 to −3.6)	.007[k]	−16.8 (−24.6 to −9.0)	.001[k]
EORTC QLQ C30[n], (score)	25.6 (14.6 to 36.5)	<.001[k]	14.6 (6.4 to 22.8)	.002[k]
LAPAQ[o] (kcal/day)	−514.7 (−866.7 to 160.7)	.008[k]	173.6 (9.5 to 337.7)	.04[k]

[a] Δ: mean difference.

[b] T0: hospital discharge.

[c] T1: 6 weeks post operation.

[d] T2: 3 months post operation.

[e] RHGS: right-hand grip strength.

[f] LHGS: left-hand grip strength.

[g] MIP: maximal inspiratory pressure.

[h] These measurements were not performed at T0 and therefore were excluded from this analysis.

[i] Missing data.

[j] 30CST: 30-second chair stand test.

[k] *P*<.05 is considered significant.

[l] 2MWT: 2-min walk test.

[m] MFI: Multidimensional Fatigue Inventory; scores range from 20 to 100, with a higher score representing more fatigue and reduced activity/motivation.

[n] EORTC QLQ C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30; scores range from 0 to 100, with high scores indicating a better quality of life.

[o] LAPAQ: Longitudinal Ageing Study Amsterdam physical activity questionnaire; total amount of activities in kilocalories per day.

Discussion

Principal Findings

To our knowledge, this is the first study demonstrating that postoperative physiotherapy with telerehabilitation is feasible in patients with postoperative complications after esophagectomy, primarily in the first 6 weeks after T0. This is in line with a study by Latham et al [11], who stated that telerehabilitation is a valuable tool to manage postoperative outcomes and functional progress directly after T0 in a patient's home environment.

The adherence rates were significantly higher in the first 6 weeks after T0 than in the following 6 weeks, where patients reported that they were generally more able to perform their ADL and were less dependent on the telerehabilitation intervention, which might explain the lower adherence rates despite a further incline in functional status. From a functional perspective, these lower adherence rates should be interpreted as a desired outcome, because it illustrates the patient's gradual independence of physiotherapeutic care.

The consistently high patient satisfaction rates of the telerehabilitation intervention in our study are confirmed in a systematic review by Mair et al [29], who stated that the greatest advantages experienced by patients were increased accessibility of specialist expertise, increased flexibility, less travel required, and reduced waiting times. This is also in agreement with the study by Moffet et al [30], who investigated patient satisfaction with in-home telerehabilitation after total knee arthroplasty and found similar results, concluding that patient satisfaction was at least equal to conventional health care delivery.

In this study, we compared patients who underwent the telerehabilitation program with a historical comparison group of patients receiving usual care and found equal functional status outcome measures at T2. This is in line with studies that found telerehabilitation interventions to be equally effective as usual care on at least one outcome measure; however, overall significant evidence in favor of telerehabilitation was still lacking [31,32].

Despite the similar functional outcomes at T2, it has to be noted that most of the functional status outcome measures of our intervention group at T0 were significantly lower than those of the matched historical comparison group. It could be argued that the intervention group gained more progress on functional status because of the physiotherapeutic treatment with telerehabilitation, in comparison with the matched historical comparison group, ultimately resulting in equal outcomes at T2.

Within the intervention group, most of the functional outcome measures significantly improved between T0-T1 and T1-T2, apart from ADL that significantly decreased during the first 6 weeks of the intervention and was restored in the following 6 weeks. A possible explanation could be that after T0, patients mostly stayed at home because they felt too weak to keep up with their ADL. Moreover, in the first 6 weeks, the telerehabilitation intervention primarily focused on increasing muscle strength of the lower extremities. After 6 weeks, the shift was gradually made toward implementing the exercises in daily life, finally resulting in a significant increase in ADL in the following 6 weeks.

Limitations

This study has intrinsic limitations. First, only 22 patients were included in this study, of which 15 patients completed the study. This might limit the generalizability of our findings. However, despite the small sample size, the included participants represented the population of interest in terms of baseline characteristics and postoperative complications. Moreover, inclusion was terminated after the inclusion of the 22nd participant because no new findings were to be expected with adding new participants to the study.

Second, this study was not a pilot feasibility trial, where patients were randomly assigned either to the intervention group or a control group to determine the effectiveness of investigational treatment. Instead, we compared the intervention group with a matched historical comparison group. Therefore, bias could not be ruled out completely.

We were not able to compare functional status outcome measures half way through the telerehabilitation intervention because the historical controls were not measured at T1.

Third, patient satisfaction was measured with a modified TSUQ that had not been validated in this specific population. Kairy et al [33] in their systematic review investigating clinical outcomes, clinical process, health care utilization, and costs associated with telerehabilitation concluded that patient satisfaction ratings were generally high, irrespective of the population. However, they also stated that operationalization and standardization of satisfaction were frequently lacking and too much focus was on the technology aspect instead of aspects of service delivery. The satisfaction questionnaire we used addressed both aspects, and therefore, we are confident that the satisfaction ratings were representative of the telerehabilitation intervention provided.

Conclusions

This study shows that patients are able to improve their functional status by doing functional exercises in their own meaningful environment supported by telerehabilitation and tablet use with distant guidance from an experienced physiotherapist. The feasibility of the physiotherapeutic intervention with telerehabilitation for this specific patient category has implications for (re)organizing postoperative physiotherapeutic care in the patient's home environment. Telerehabilitation cannot replace face-to-face physiotherapy as physical examination remains to be necessary, but taking into account positive adherence rates and satisfaction, we strongly suggest considering this way of treatment delivery for patients with esophageal cancer treated with surgery and having postoperative complications, especially in the first 6 weeks after T0. We also recommend investigating the potential cost-effectiveness of telerehabilitation compared with usual care. Although we found equal functional status outcomes in both the intervention group and the historical comparison group at T2, we suggest performing a randomized controlled trial to draw firm conclusions on its effectiveness.

Acknowledgments

The authors would like to acknowledge Sander Steenhuizen, PT, and Sven Geelen, PT, MSc, for recruiting participants and performing measurements. The authors would also like to acknowledge Physitrack, who provided their services for this research project free of charge without being involved in the design, execution, analysis, and conclusions of this research. This work was supported by a Doctoral Grant for Teachers (023.003.016) in 2014 by the Dutch Research Council (Nederlandse organisatie voor Wetenschappelijk Onderzoek). MB reports research grants from Olympus and Stryker, in addition to consulting fees from Medtronic, Johnson and Johnson, and Mylan, in areas unrelated to this study.

Conflicts of Interest

None declared.

References

1. Grotenhuis BA, Wijnhoven BP, Grüne F, van Bommel J, Tilanus HW, van Lanschot JJ. Preoperative risk assessment and prevention of complications in patients with esophageal cancer. J Surg Oncol 2010 Mar 1;101(3):270-278. [doi: 10.1002/jso.21471] [Medline: 20082349]
2. Feeney C, Reynolds JV, Hussey J. Preoperative physical activity levels and postoperative pulmonary complications post-esophagectomy. Dis Esophagus 2011 Sep;24(7):489-494. [doi: 10.1111/j.1442-2050.2010.01171.x] [Medline: 21309920]
3. O'Neill L, Moran J, Guinan EM, Reynolds JV, Hussey J. Physical decline and its implications in the management of oesophageal and gastric cancer: a systematic review. J Cancer Surviv 2018 Aug;12(4):601-618. [doi: 10.1007/s11764-018-0696-6] [Medline: 29796931]

4. Dronkers JJ, Lamberts H, Reutelingsperger IM, Naber RH, Dronkers-Landman CM, Veldman A, et al. Preoperative therapeutic programme for elderly patients scheduled for elective abdominal oncological surgery: a randomized controlled pilot study. Clin Rehabil 2010 Jul;24(7):614-622. [doi: 10.1177/0269215509358941] [Medline: 20530651]
5. Hulzebos EH, Helders PJ, Favié NJ, de Bie RA, Brutel de la Riviere A, van Meeteren NL. Preoperative intensive inspiratory muscle training to prevent postoperative pulmonary complications in high-risk patients undergoing CABG surgery: a randomized clinical trial. J Am Med Assoc 2006 Oct 18;296(15):1851-1857. [doi: 10.1001/jama.296.15.1851] [Medline: 17047215]
6. van Egmond MA, van der Schaaf M, Klinkenbijl JH, Engelbert RH, van Berge Henegouwen MI. Preoperative functional status is not associated with postoperative surgical complications in low risk patients undergoing esophagectomy. Dis Esophagus 2017 Jan 1;30(1):1-7. [doi: 10.1111/dote.12469] [Medline: 26918788]
7. Dettling DS, van der Schaaf M, Blom RL, Nollet F, Busch OR, van Berge Henegouwen MI. Feasibility and effectiveness of pre-operative inspiratory muscle training in patients undergoing oesophagectomy: a pilot study. Physiother Res Int 2013 Mar;18(1):16-26. [doi: 10.1002/pri.1524] [Medline: 22489016]
8. Santa Mina D, Alibhai SM, Matthew AG, Guglietti CL, Steele J, Trachtenberg J, et al. Exercise in clinical cancer care: a call to action and program development description. Curr Oncol 2012 Jun;19(3):e136-e144 [FREE Full text] [doi: 10.3747/co.19.912] [Medline: 22670103]
9. Guinan EM, Dowds J, Donohoe C, Reynolds JV, Hussey J. The physiotherapist and the esophageal cancer patient: from prehabilitation to rehabilitation. Dis Esophagus 2017 Jan 1;30(1):1-12. [doi: 10.1111/dote.12514] [Medline: 27862675]
10. Laver KE, Schoene D, Crotty M, George S, Lannin NA, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2013 Dec 16(12):CD010255 [FREE Full text] [doi: 10.1002/14651858.CD010255.pub2] [Medline: 24338496]
11. Latham NK, Harris BA, Bean JF, Heeren T, Goodyear C, Zawacki S, et al. Effect of a home-based exercise program on functional recovery following rehabilitation after hip fracture: a randomized clinical trial. J Am Med Assoc 2014 Feb 19;311(7):700-708 [FREE Full text] [doi: 10.1001/jama.2014.469] [Medline: 24549550]
12. van Egmond MA, van der Schaaf M, Vredeveld T, Vollenbroek-Hutten MM, van Berge Henegouwen MI, Klinkenbijl JH, et al. Effectiveness of physiotherapy with telerehabilitation in surgical patients: a systematic review and meta-analysis. Physiotherapy 2018 Sep;104(3):277-298. [doi: 10.1016/j.physio.2018.04.004] [Medline: 30030037]
13. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med 2003 Feb 4;138(3):161-167. [doi: 10.7326/0003-4819-138-3-200302040-00007] [Medline: 12558354]
14. Head BA, Studts JL, Bumpous JM, Gregg JL, Wilson L, Keeney C, et al. Development of a telehealth intervention for head and neck cancer patients. Telemed J E Health 2009 Jan;15(1):44-52 [FREE Full text] [doi: 10.1089/tmj.2008.0061] [Medline: 19199847]
15. Thompson DA, Leimig R, Gower G, Winsett RP. Assessment of depressive symptoms during post-transplant follow-up care performed via telehealth. Telemed J E Health 2009 Sep;15(7):700-706 [FREE Full text] [doi: 10.1089/tmj.2009.0021] [Medline: 19694599]
16. Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg 2009 Aug;250(2):187-196. [doi: 10.1097/SLA.0b013e3181b13ca2] [Medline: 19638912]
17. Stevens A, Beurskens A, Köke A, van der Weijden T. The use of patient-specific measurement instruments in the process of goal-setting: a systematic review of available instruments and their feasibility. Clin Rehabil 2013 Nov;27(11):1005-1019. [doi: 10.1177/0269215513490178] [Medline: 23881336]
18. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
19. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]
20. Lorenz DS, Reiman MP, Walker JC. Periodization: current review and suggested implementation for athletic rehabilitation. Sports Health 2010 Nov;2(6):509-518 [FREE Full text] [doi: 10.1177/1941738110375910] [Medline: 23015982]
21. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]
22. World Health Organization. International Classification of Functioning, Disability and Health: ICF. Geneva: World Health Organization; 2001.
23. Trutschnigg B, Kilgour RD, Reinglas J, Rosenthall L, Hornby L, Morais JA, et al. Precision and reliability of strength (Jamar vs. Biodex handgrip) and body composition (dual-energy X-ray absorptiometry vs. bioimpedance analysis) measurements in advanced cancer patients. Appl Physiol Nutr Metab 2008 Dec;33(6):1232-1239. [doi: 10.1139/H08-122] [Medline: 19088782]

24. Timmerman H, de Groot JF, Hulzebos HJ, de Knikker R, Kerkkamp HE, van Meeteren NL. Feasibility and preliminary effectiveness of preoperative therapeutic exercise in patients with cancer: a pragmatic study. Physiother Theory Pract 2011 Feb;27(2):117-124. [doi: 10.3109/09593981003761509] [Medline: 20690877]
25. Millor N, Lecumberri P, Gómez M, Martínez-Ramírez A, Izquierdo M. An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit. J Neuroeng Rehabil 2013 Aug 1;10:86. [doi: 10.1186/1743-0003-10-86] [Medline: 24059755]
26. Brooks D, Parsons J, Tran D, Jeng B, Gorczyca B, Newton J, et al. The two-minute walk test as a measure of functional capacity in cardiac surgery patients. Arch Phys Med Rehabil 2004 Sep;85(9):1525-1530. [doi: 10.1016/j.apmr.2004.01.023] [Medline: 15375829]
27. Smets EM, Garssen B, Bonke B, de Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995 Apr;39(3):315-325. [doi: 10.1016/0022-3999(94)00125-o] [Medline: 7636775]
28. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 3;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
29. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. Br Med J 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [doi: 10.1136/bmj.320.7248.1517] [Medline: 10834899]
30. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. Patient satisfaction with in-home telerehabilitation after total knee arthroplasty: Results from a randomized controlled trial. Telemed J E Health 2017 Feb;23(2):80-87. [doi: 10.1089/tmj.2016.0060] [Medline: 27529575]
31. Hawkes AL, Gollschewski S, Lynch BM, Chambers S. A telephone-delivered lifestyle intervention for colorectal cancer survivors 'CanChange': a pilot study. Psychooncology 2009 Apr;18(4):449-455. [doi: 10.1002/pon.1527] [Medline: 19241477]
32. van den Brink JL, Moorman PW, de Boer MF, Hop WC, Pruyn JF, Verwoerd CD, et al. Impact on quality of life of a telemedicine system supporting head and neck cancer patients: a controlled trial during the postoperative period at home. J Am Med Inform Assoc 2007;14(2):198-205 [FREE Full text] [doi: 10.1197/jamia.M2199] [Medline: 17213498]
33. Kairy D, Lehoux P, Vincent C, Visintin M. A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation. Disabil Rehabil 2009;31(6):427-447. [doi: 10.1080/09638280802062553] [Medline: 18720118]

Abbreviations

2MWT: 2-min walk test
30CST: 30-second chair stand test
ADL: activities of daily life
eHealth: electronic health
HRQL: health-related quality of life
LAPAQ: Longitudinal Ageing Study Amsterdam Physical Activity Questionnaire
LoS: length of hospital stay
METC: Medical Ethical Committee
TSUQ: telemedicine satisfaction and usefulness questionnaire

Edited by G Eysenbach; submitted 30.08.19; peer-reviewed by M Vollenbroek-Hutten, H Friess, A Van de Winckel; comments to author 09.12.19; revised version received 22.01.20; accepted 07.02.20; published 09.06.20.

Please cite as:
van Egmond MA, Engelbert RHH, Klinkenbijl JHG, van Berge Henegouwen MI, van der Schaaf M
Physiotherapy With Telerehabilitation in Patients With Complicated Postoperative Recovery After Esophageal Cancer Surgery: Feasibility Study
J Med Internet Res 2020;22(6):e16056
URL: http://www.jmir.org/2020/6/e16056/
doi:10.2196/16056
PMID:32515742

Original Paper

Evaluating Safety and Efficacy of Follow-up for Patients With Abdominal Pain Using Video Consultation (SAVED Study): Randomized Controlled Trial

Dinesh Visva Gunasekeran[1,2*], MBBS; Zhenghong Liu[1*], MBBS; Win Jim Tan[1], MBBS; Joshua Koh[1], MBBS; Chiu Peng Cheong[1]; Lay Hong Tan[1]; Chee Siang Lau[1]; Gaik Kheng Phuah[1]; Newsie Donnah A Manuel[1]; Che Chong Chia[1]; Gek Siang Seng[1]; Nancy Tong[1]; May Hang Huin[1]; Suzette Villaluna Dulce[1]; Susan Yap[1]; Kishanti Ponampalam[1]; Hao Ying[3], PhD; Marcus Eng Hock Ong[1,3,4], MBBS, FRCSE; R Ponampalam[1], MBBS, FRCSE

[1]Department of Emergency Medicine, Singapore General Hospital, Singapore, Singapore

[2]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[3]Health Services Research Center, Singhealth Services, Singapore, Singapore

[4]Health Services and Systems Research, Duke–National University of Singapore Medical School, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
R Ponampalam, MBBS, FRCSE
Department of Emergency Medicine
Singapore General Hospital
Singapore
Singapore
Phone: 65 6321 3497
Email: ponampalam@singhealth.com.sg

Abstract

Background: The benefits of telemedicine include cost savings and decentralized care. Video consultation is one form that enables early detection of deteriorating patients and promotion of self-efficacy in patients who are well but anxious. Abdominal pain is a common symptom presented by patients in emergency departments. These patients could benefit from video consultation, as it enables remote follow-up of patients who do not require admission and facilitates early discharge of patients from overcrowded hospitals.

Objective: The study aimed to evaluate the safety and efficacy of the use of digital telereview in patients presenting with undifferentiated acute abdominal pain.

Methods: The SAVED study was a prospective randomized controlled trial in which follow-up using existing telephone-based telereview (control) was compared with digital telereview (intervention). Patients with undifferentiated acute abdominal pain discharged from the emergency department observation ward were studied based on intention-to-treat. The control arm received routine, provider-scheduled telereview with missed reviews actively coordinated and rescheduled by emergency department staff. The intervention arm received access to a platform for digital telereview (asynchronous and synchronous format) that enabled patient-led appointment rescheduling. Patients were followed-up for 2 weeks for outcomes of service utilization, efficacy (compliance with their disposition plan), and safety (re-presentation for the same condition).

Results: A total of 70 patients participated, with patients randomly assigned to each arm (1:1 ratio). Patients were a mean age of 40.0 (SD 13.8; range 22-71) years, predominantly female (47/70, 67%), and predominantly of Chinese ethnicity (39/70, 56%). The telereview service was used by 32 patients in the control arm (32/35, 91%) and 18 patients in the intervention arm (18/35, 51%). Most patients in control (33/35, 94%; 95% CI 79.5%-99.0%) and intervention (34/35, 97%; 95% CI 83.4%-99.9%) arms were compliant with their final disposition. There was a low rate of re-presentation at 72 hours and 2 weeks for both control (72 hours: 2/35, 6%; 95% CI 1.0%-20.5%; 2 weeks: 2/35, 6%, 95% CI 1.0%-20.5%) and intervention (72 hours: 2/35, 6%; 95% CI 1.0%-20.5%; 2 weeks: 3/35, 9%, 95% CI 2.2%-24.2%) arms. There were no significant differences in safety (P>.99) and efficacy (P>.99) between the two groups.

Conclusions: The application of digital telereview for the follow-up of patients with abdominal pain may be safe and effective. Future studies are needed to evaluate its cost-effectiveness and usefulness for broader clinical application.

Trial Registration: ISRCTN Registry ISRCTN28468556; http://www.isrctn.com/ISRCTN28468556.

(J Med Internet Res 2020;22(6):e17417) doi:10.2196/17417

KEYWORDS

digital health; teleconsultation; video consultation; telereview; abdominal pain; primary care; emergency department

Introduction

Existing literature on telemedicine suggests that it enables cost savings and improved health care access for patients with diverse illnesses [1,2]. These reports have fueled a rise in the adoption of telemedicine for these applications in various clinical settings [3], allowing new models for decentralized care that may help alleviate shortages in health care resources and help encourage self-management by patients where appropriate [4]. This need has been heightened with the coronavirus disease 2019 (COVID-19) pandemic which has been placing many health systems in dire straits. The application of video consultation to facilitate early discharge and subsequent remote follow-up of relatively well patients is a potential solution for increasingly oversubscribed emergency departments [5], given the detrimental effects of overcrowding on the timeliness and the quality of care [6].

Patient and provider acceptance of video consultation has grown in recent years, and over two-thirds of respondents to a survey of patient and caregiver acceptance conducted in the emergency department of Singapore General Hospital were comfortable using mobile technology to share information [7]. Prior studies in Singapore have reported benefits from telemedicine such as improved health care access in the form of remote consultation mediated by health professionals for acute illnesses such as poisoning [8] or maritime emergencies [9]; however, few international studies have reported controlled outcomes from the use of telemedicine for acute illnesses by patients directly. Furthermore, many studies had considerable limitations in terms of study design [10] and clarity in reporting which telemedicine interventions were used [11].

The use of video consultation for telereview has not been conclusively investigated in patients with acute gastrointestinal ailments [12,13]. The objective of this study was to evaluate the safety and efficacy of digital telereview for patients in the emergency department who present with undifferentiated acute abdominal pain. In this study, digital telereview was evaluated as a “pull-from-patient”, a patient-led form of service delivery, and was compared with existing telephone-based telereview evaluated as a “push-to-patient”, a provider-led form of service delivery. This was the first direct and pragmatic trial of these alternative forms of service delivery for the follow-up of patients with abdominal pain.

Methods

Study Design

The SAVED (Safety and Efficacy of Follow-up for Patients With Abdominal Pain Using Video Consultation) study was a prospective randomized controlled trial with a 1:1 allocation ratio. An existing telephone-based telereview service (control) was compared with digital telereview using DoctorBell—a novel, web-based telehealth platform (intervention). This study was approved by the Centralized Institutional Review Board of SingHealth Singapore General Hospital (protocol number 2017/2049) and conducted in accordance with the Declaration of Helsinki (2000).

Study Setting

The emergency department was selected as the location for this pilot study because 24-hour medical services were available to attend to any patients whose condition deteriorated [13]. Patients from the emergency department observation ward with severe undifferentiated illnesses that did not meet clinical criteria for hospital admission were recruited. This cohort represents a segment of the population of patients in the emergency department for whom the decision (final disposition) is delayed by clinical uncertainty [14,15]; therefore, the recruitment of patients discharged from the emergency department observation ward allowed investigators to stress-test the appropriateness of disposition assigned by digital telereview as well as patient adherence to recommendations, since telephone-based telereview is routinely conducted within 48 to 72 hours of discharge for patients with undifferentiated illness as a fail-safe measure that enables early detection of patients with deteriorating clinical condition. This practice improves care continuity while enabling safe discharge to address overcrowding.

Study Recruitment

Patients with undifferentiated acute abdominal pain who were discharged from a tertiary hospital in Singapore using detailed criteria (Abdominal Pain or Gastroenteritis Pathways in Multimedia Appendix 1) were considered for recruitment at the point of discharge by study team members after the patient had received routine treatment and a disposition plan which included counselling for self-efficacy and monitoring at home as well as education regarding clinical features that warrant a return to the emergency department (re-presentation).

From September 2017 to May 2018, consecutive patients who presented to the emergency department and who satisfied the study criteria were recruited. Inclusion criteria were the ability to read in English and the ability to operate smartphone messaging apps (such as WhatsApp). To exclude patients who may be considered vulnerable, exclusion criteria were defined as less than 21 years of age, pregnant, a prisoner, cognitively impaired, or requiring a legal representative for informed consent in Singapore. As a pilot, formal sample size calculation was not possible due to a lack of relevant data.

Randomization and Masking

After completion of informed consent, participants were randomly assigned by a study team member who withdrew lots from a box containing equal numbers of paper indicating either control or intervention. Lots in the box were replenished after each draw by a study team member (S.Y.) who was not involved in clinical care or the implementation of the randomization. At recruitment, study team members provided participants with links to web-based surveys about symptoms to be reported by patients at initial presentation and at follow-up, following recruitment and telereview, respectively. Blinding was not possible in this study.

Intervention

Follow-up after discharge by telereview is routinely conducted by triage nurses to facilitate early discharge from the observation ward, with follow-up review to ascertain right-siting through prompt identification of deteriorating patients as well as encouraging self-efficacy in well patients who are anxious but who do not need to re-visit the emergency department. During telereview, patients who are well are advised on self-management while patients with ominous symptomatology or deteriorating illness are advised to return to the emergency department for further evaluation. In this study, patients in the control arm received routine, provider-led telephone-based telereview after discharge. Telereview was conducted by the emergency department staff on-shift at the time of the patient's booked appointment. Any missed telereviews were actively and manually rescheduled by staff with three attempted phone calls within 48 to 72 hours following discharge. No further attempt was made to contact patients who were not reached in the 72-hour period or those who declined telereview.

Patients in the intervention arm had access to DoctorBell, a novel telehealth platform accessible on smartphone or desktop by web browsers. This was designed using a design-thinking process based on the context and workflows of an emergency department. It allowed patient-led booking, rescheduling, or cancellation of one digital telereview appointment based on the patient's own individual availability, restricted to 48- to 72-hour window following discharge from the emergency department. Digital telereview appointments through this platform were patient-led and were not actively rescheduled by emergency department staff if missed or cancelled. Before the digital telereview appointment (synchronous teleconsultation with video, voice, and text messaging), patients received an in-app form that allowed them to report important history or symptoms beforehand (asynchronous). This was sent to the emergency department staff on-duty before video consultation, providing them with the opportunity to clarify any uncertainties with the on-duty attending physician before synchronous consultation. Upon starting work, staff could view digital telereview bookings scheduled during their shift and received automated real-time push notifications of any changes made to appointments.

Study Outcomes and Statistical Analysis

Patients in both groups were followed up for 2 weeks to examine the study outcomes of service utilization (telereview appointments used), efficacy (patient compliance with their final disposition plan), and safety (re-presentation to the emergency department for the same medical problem within 72 hours or within 2 weeks). Initially, service utilization and safety were the only outcome measures planned for investigation; however, study team members observed that not all patients complied with instructions given during telereview to return to the emergency department. Therefore, efficacy was later added to the analysis as an outcome measure using existing data. No change to study procedures was required. For patients who did not receive telereview, the final disposition plan was that given at the point of discharge from the emergency department (ie, self-management and monitoring at home). For patients who received telereview, the final disposition plan was that given during telereview (ie, whether to continue self-management or return to the emergency department).

The study population was analyzed based on intention-to-treat. All hypotheses were two-sided with a $P<.05$ considered statistically significant. Descriptive statistical analysis was conducted using SPSS software (version 20.0; IBM Corp). Associations between categorical variables were analyzed using Fisher exact test. Where expected counts within all categories were greater than 5, chi-square test was used instead. Associations between continuous variables were analyzed using the two-tailed two-sample *t* test.

Results

From September 2017 to May 2018, patients (N=72) who were discharged from the emergency department observation ward and who satisfied inclusion criteria were recruited to participate in this study. One patient was under the age of 21 and was excluded. Another patient declined to participate in the study. The remaining patients (N=70) were enrolled with patients (n=35) randomized to each arm (CONSORT diagram, Figure 1). One patient crossed over from digital telereview to telephone-based telereview since, rather than booking a digital telereview appointment, the patient called the emergency department within 72 hours to report persistent symptoms and complied with instructions to return to the emergency department the next day (on the fourth day postdischarge). This patient was analyzed in the intervention arm in accordance with intention-to-treat.

There were no significant demographic differences between the study groups. Demographics of the study population are described in detail in Table 1. Patients were a mean of 40.0 (SD 13.8; range 22-71) years of age, mostly female (47/70, 67%) and mostly of Chinese ethnicity (39/70, 56%). Symptoms reported at initial presentation are detailed in Table 2 and those reported at follow-up are detailed in Table 3 demonstrating the usefulness of a structured web survey in gathering symptomatology during telereview.

Figure 1. A CONSORT diagram depicting the study. Patients in the control arm receive the push-to-patient form of telephone telereview while patients in the intervention arm receive the pull-from-patient form of digital telereview.

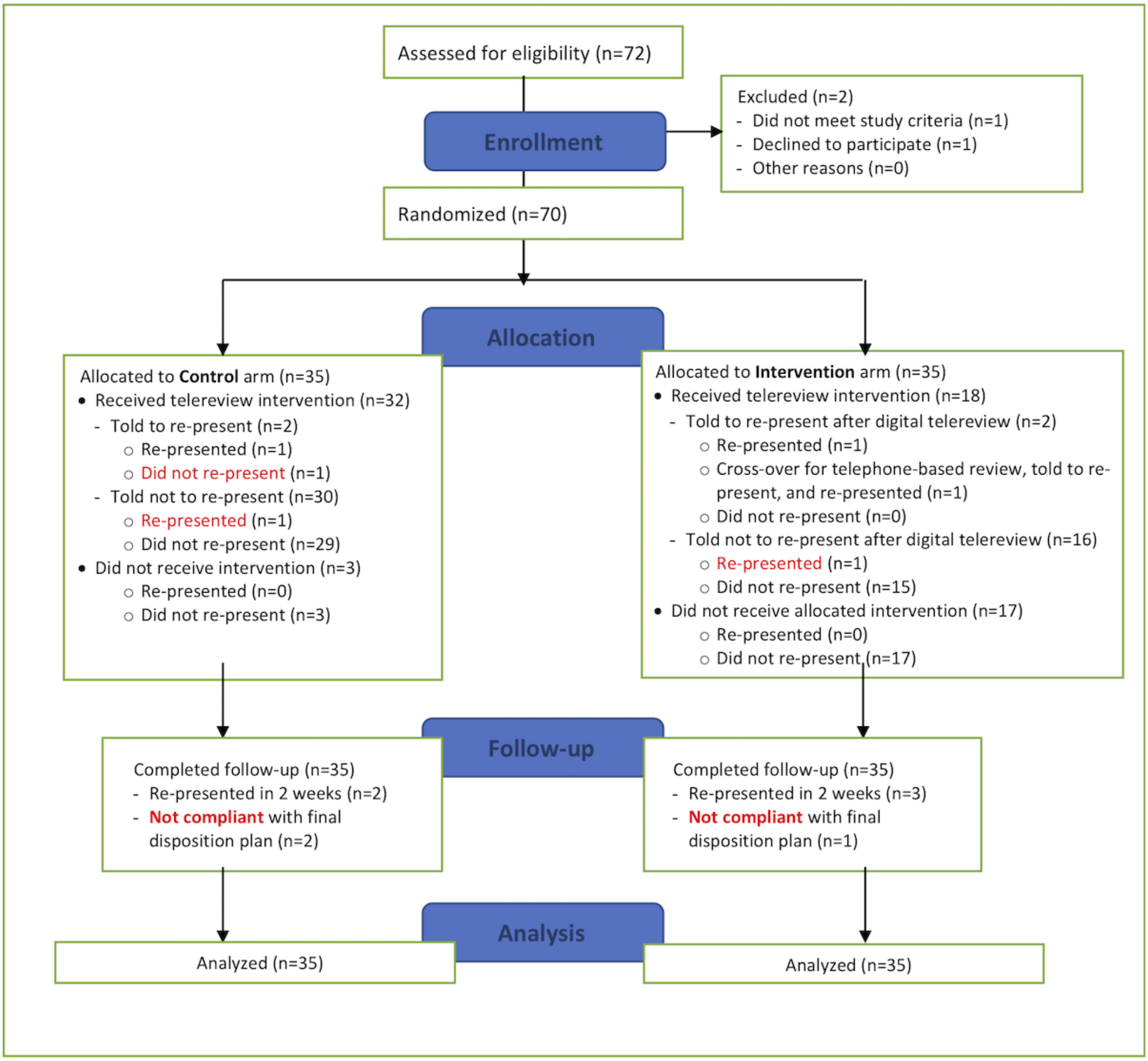


Table 1. Patient Demographics.

Variable	All (N=70)	DoctorBell (n=35)	Telephone (n=35)	Chi-square (*df*)	*P* value
Age (years)				N/A[a]	.88[b]
mean (SD)	40.0 (13.8)	40.2 (13.3)	39.7 (14.4)		
range	22-71	26-69	22-71		
Gender, n (%)				0 (1)	>.99
Male	23 (33)	12 (34)	11 (31)		
Female	47 (67)	23 (66)	24 (69)		
Ethnicity, n (%)				N/A	0.38
Chinese	39 (56)	18 (51)	21 (60)		
Malay	16 (23)	7 (20)	9 (26)		
Indian	6 (9)	3 (9)	3 (9)		
Other	9 (13)	7 (20)	2 (6)		
Nationality, n (%)				N/A	0.48
Singaporean/Permanent resident	61 (87)	29 (83)	32 (91)		
Other	9 (13)	6 (17)	3 (9)		
Pre-study survey respondent, n (%)[c]				N/A	>.99
Patient	56 (92)	28 (93)	28 (90)		
Caregiver	3 (5)	1 (3)	2 (6)		
Both	2 (3)	1 (3)	1 (3)		
Disease manager, n (%)[c]				N/A	.35
Patient	56 (92)	29 (97)	27 (87)		
Caregiver	0 (0)	0 (0)	0 (0)		
Both	5 (8)	1 (3)	4 (13)		
Patient highest level of education, n (%)[c]				N/A	.38
None	1 (2)	0 (0)	1 (3)		
Primary	2 (3)	2 (7)	0 (0)		
Secondary	8 (13)	2 (7)	6 (19)		
Postsecondary Diploma/Certificate	19 (31)	11 (37)	8 (26)		
Degree	24 (39)	12 (40)	12 (39)		
Masters/PhD	7 (11)	3 (10)	4 (13)		
Patient needed relative to accompany to hospital, n (%)[c]				0.8 (1)	.37
Yes	37 (61)	16 (53)	21 (68)		
No	24 (39)	14 (47)	10 (32)		

[a] N/A: not applicable as Fisher's exact test was used.

[b] A two-sided *t* test was used here.

[c] This number is less than the group number because respondents did not submit their surveys (n=9, n=5, and n=4 missing in All, DoctorBell, and Telephone, respectively).

Table 2. Patient symptoms at presentation to hospital reported in a web-based survey via a link provided following patient recruitment.

Variable	All (n=61)[a]	DoctorBell (n=30)[a]	Telephone (n=31)[a]	Chi-square (*df*)	*P* value
Previous abdominal surgery	14 (23)	5 (17)	9 (29)	0.7 (1)	.40
Abdominal bloating	39 (64)	20 (67)	19 (61)	0 (1)	.87
Loss of appetite	50 (82)	23 (77)	27 (87)	0.5 (1)	.47
Fever	16 (26)	9 (30)	7 (23)	0.1 (1)	.71
Giddiness	34 (56)	16 (53)	18 (58)	0 (1)	.91
Blood in stools	2 (3)	1 (3)	1 (3)	N/A[b]	>.99
Malena	1 (2)	1 (3)	0 (0)	N/A	.49
Diarrhea	35 (57)	16 (53)	19 (61)	0.1 (1)	.71
Pale stools	7 (11)	4 (13)	3 (10)	N/A	.71
Nausea/Vomiting	42 (69)	24 (80)	18 (58)	2.5 (1)	.12

[a] This number is less than the group number because respondents did not submit their surveys (n=9, n=5, and n=4 missing in All, DoctorBell, and Telephone, respectively).

[b] N/A: not applicable as Fisher's exact test was used.

Table 3. Patient symptoms on follow-up teleconsultation reported in a web-based survey via a link provided following patient recruitment.

Variable	All (n=43)[a]	DoctorBell (n=14)[a]	Telephone (n=29)[a]	Chi-square (*df*)	*P* value
Abdominal pain	12 (28)	3 (21)	9 (31)	N/A[b]	.72
Abdominal bloating	8 (19)	3 (21)	5 (17)	N/A	>.99
Diarrhea	12 (28)	3 (21)	9 (31)	N/A	.72
Nausea/vomiting	3 (7)	1 (7)	2 (7)	N/A	>.99
Giddiness	1 (2)	0 (0)	1 (3)	N/A	>.99

[a] This number is less than the group number because respondents did not submit their surveys (n=9, n=5, and n=4 missing in All, DoctorBell, and Telephone, respectively).

[b] N/A: not applicable as Fisher's exact test was used.

There was an overall 71% (50/70) utilization of the telereview service by 32 patients in the telephone-based telereview control arm (32/35, 91%) and 18 patients in the digital telereview intervention arm (18/35, 51%). Most patients in control (33/35, 94%; 95% CI 79.5%-99.0%) and intervention (34/35, 97%; 95% CI 83.4%-99.9%) arms were compliant with final disposition. There was a low rate of re-presentation at 72 hours for both arms (control: 2/35, 6%; 95% CI 1.0%-20.5%; intervention: 2/35, 6%; 95% CI 1.0%-20.5%) and at 2 weeks for both control (2/35, 6%; 95% CI 1.0%-20.5%) and intervention (3/35, 9%; 95% CI 2.2%-24.2%) arms. After the initial 72-hour period, no patients represented within the control arm crossed over whereas the single patient from the intervention arm who crossed over to telephone-based review re-presented at 4 days. Using the chi-square test, there were no significant differences between the control and intervention arms with regards to efficacy as well as safety in terms of re-presentation within 72 hours and within 2 weeks (*P*>.99).

Discussion

Principal Findings

Emergency department overcrowding is a persistent challenge despite increased human resource capacity and process innovations [16] including emergency department observation wards themselves [15]. This has been exacerbated by the COVID-19 pandemic with front-loaded emergency services. New technologies may help resolve overcrowding through automated solutions such as machine learning to optimize existing processes [17] or digital telemedicine to enable new processes that streamline the flow of patients [3]. This study is the first pragmatic randomized controlled trial that evaluated efficacy of digital telereview for the follow-up of patients with undifferentiated acute abdominal pain. This solution uses a hybrid of asynchronous and synchronous teleconsultation—the unique benefits of which have been outlined in a review and case study of their application in eye care [18]. This study is a timely contribution to ongoing debate surrounding the effective implementation of remote consultation in tertiary care. The results of this study suggest that digital telereview may be a safe and effective tool to optimize follow-up processes for right-siting patients in tertiary care.

Digital platforms may enable safe re-design of existing processes from a push-to-patient form of service delivery to a pull-from-patient form. Earlier studies have indicated that patients may not always have adequate insight about the severity of their illness [19]. Hildebrandt et al [19] surveyed primary care physicians regarding patient-directed afterhours telephone triage services. Calls from patients who considered their conditions to be non-emergent were reviewed by physicians who found that roughly half of these patients (range 22%-77%) required urgent review [19]. In this study, telereview following initial physician contact was selected instead of teletriage in order to ensure that all patients were first counselled by a health professional regarding their condition before intervention.

The use of telereview may not be appropriate in the presence of certain clinical factors. These include patient factors such as lack of insight regarding their health or lack of familiarity with the use of mobile phone technology [7,20]. Some researchers posit that dedicated staff should be rostered to respond to any urgent enquiries and actively follow-up all patients to ensure timely assessment [19]; however, the results from this study indicate that patient-led digital telereview following routine education and discharge advice can be as effective as provider-led active follow-up despite the finding of reduced resource utilization (patient-led digital telereview: 18/35, 51%; provider-led telephone-based telereview: 32/35, 91%; Figure 1).

This pull-from-patient service delivery reduces manpower utilization to coordinate unnecessary telereview in patients who are well. Findings from a retrospective study of 1522 video-conference teletriage interventions over 52 weeks support this notion that teleconsultation can improve clinical efficiency [21]. Similarly, Brennan et al [22] reported good outcomes and time saved from emergency department teleconsultations. In this study, the digital platform served as a filter to triage patients, saved time for emergency department staff by collecting fundamental patient-reported history asynchronously, and presented the information accessibly to emergency department staff for clarification with the attending physician at any suitable time before synchronous consultation (instead of traditionally having to place the patient on-hold and interrupt the attending for advice). The results from the current study will facilitate the necessary health economic assessments to guide the implementation of digital telereview, which need to be conducted based on relevant outcomes that have been identified in the literature such as reduced adverse events and reduced time requirement [23].

Limitations

Limitations of this study include a lack of pre-existing data for sample size calculation and possible selection bias since patients were only recruited when study team members were on shift—a common constraint of studies that are conducted in busy emergency departments [13]. Furthermore, these results may not be generalizable given that participants were relatively young and well-educated although, from a practical perspective, such patients are more likely to use digital consultation services [7,23]. Other limitations are consistent with intention-to-treat analysis such as conservative estimates for efficacy as a result of totals diluted by patients who did not receive the intervention (Figure 1). The strengths of intention-to-treat analysis are that it evaluates the actual performance compared objectively with existing practice by considering any deviations in protocol. Further strengths of this study are the use of the gold standard randomized controlled methodology and the comprehensive description of interventions, participants, and outcomes.

Existing reports caution that successful implementation of a telemedicine intervention may not necessarily be reproduced once the clinical application is even slightly modified. This has been observed with investigations of the WelTel telehealth intervention [2]. Lester et al [2] first reported successful implementation of WelTel for messaging that provides weekly automated notifications for patients with *Human immunodeficiency virus* (and promotes adherence to anti-retroviral therapy with messages based on patient-reported adherence. When van der Kop et al [24] evaluated use of WelTel to promote attendance to follow-up appointments, they found lower compliance in the patients who received the intervention and no significant difference between the groups. Therefore, the results from our study may not be generalizable to other undifferentiated acute presentations such as chest pain. Physicians, administrators, and researchers should be mindful of factors such as clinical context and the form of intervention in the consideration of existing evidence to guide telehealth implementation [20].

The benefits of applying design-thinking in this manner have been described for numerous telehealth applications including tele-ophthalmology enabled by artificial intelligence since contextual considerations such as lack of adequate infrastructure or stable internet access may have an impact on model implementation [25]. In the context of the COVID-19 pandemic, there is a heightened need for providers to evaluate models such as these to rapidly scale health system capacity to address clinical need and to decentralize services to reduce health care–associated transmission [26]. The concurrent use of asynchronous modalities as a filter before synchronous consultation may help ensure better allocation of healthcare manpower similar to the model in this study. Looking beyond the pandemic, when structured appropriately to address the clinical need, digital telereview developed using a design that is mindful of clinical context to guide telehealth implementation (as described above) may lead to considerable cost savings (as suggested by our results). This has already been definitively demonstrated at scale for other related technology applications such as synchronous artificial intelligence–enabled tele-ophthalmology [27] and asynchronous multicenter cloud-computing enabled registry-based research [28].

Conclusions

Digital telereview may safely facilitate a re-design from a push-to-patient to a pull-from-patient form of follow-up in select patients. This study suggests that there is potential to save cost and manpower-time with digital telereview and has laid the ground work for future investigations to examine the benefits of implementing digital telereview at scale for various clinical applications.

Acknowledgments

Patrick Poh Huan Yu, Ming Xiang Tham, and Nicholas Chew Yi Ming are acknowledged as clinical innovators and research collaborators for developing the novel DoctorBell telehealth platform used in this study. The DoctorBell platform was acquired by Mobile Health Pte Ltd on August 1, 2018 and integrated into the MaNaDr telehealth platform.

DVG received travel funding from the Mobile Health Education grant, Commonwealth Fellowship in Innovation award, and the National Youth Fund award for clinical research training and collaborations at Oxford University and Stanford University. The study was partially funded by the Singapore General Hospital in the form of the research hours contributed by study team members employed by the hospital. Funder(s) played no part in the design, conduct, or decision to publish for this study.

This manuscript is dedicated to the memory of our colleague and friend, the late SY, who succumbed to her illness in the months following the completion of this study. She was a dedicated researcher without whom this study among many others would not have been possible. Her passing is a profound loss and her contributions to medicine and clinical sciences will live on.

Conflicts of Interest

DVG reports investment in DoctorBell, VISRE, AskDr and Shyfts. He reports serving as lecturer and faculty advisor to the medical innovation program of the National University of Singapore, as well as Head of Health Informatics (Collaborative Ocular Tuberculosis Study group).

Multimedia Appendix 1
Singapore General Hospital (SGH) Emergency Observation Ward Gastroenteritis and Abdominal Pain protocols.
[PDF File (Adobe PDF File), 447 KB - jmir_v22i6e17417_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1276 KB - jmir_v22i6e17417_app2.pdf]

References

1. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/nejmsr1503323]
2. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. The Lancet 2010 Nov;376(9755):1838-1845. [doi: 10.1016/s0140-6736(10)61997-6]
3. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
4. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
5. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. PLoS One 2018 Aug 30;13(8):e0203316 [FREE Full text] [doi: 10.1371/journal.pone.0203316] [Medline: 30161242]
6. Lowthian JA, Curtis AJ, Cameron PA, Stoelwinder JU, Cooke MW, McNeil JJ. Systematic review of trends in emergency department attendances: an Australian perspective. Emerg Med J 2011 May 20;28(5):373-377. [doi: 10.1136/emj.2010.099226] [Medline: 20961936]
7. Liu Z, Ng M, Gunasekeran DV, Li H, Ponampalam K, Ponampalam R. Mobile technology: Usage and perspective of patients and caregivers presenting to a tertiary care emergency department. World J Emerg Med 2020;11(1):5-11 [FREE Full text] [doi: 10.5847/wjem.j.1920-8642.2020.01.001] [Medline: 31892997]
8. Ponampalam R, Loh C. Cost Benefits of the Drug and Poison Information Center (DPIC) in preventing unnecessary hospitalization: The Singapore Experience. In. Vol 17: Hong Kong Journal of Emerg Med; 2010:45-53.
9. Lateef F, Anantharaman V. Maritime radio-medical services: the Singapore General Hospital experience. Am J Emerg Med 2002 Jul;20(4):349-351. [doi: 10.1053/ajem.2002.33782] [Medline: 12098185]
10. Mold F, Hendy J, Lai Y, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
11. Zanaboni P, Lettieri E. Institutionalizing telemedicine applications: the challenge of legitimizing decision-making. J Med Internet Res 2011 Sep 28;13(3):e72 [FREE Full text] [doi: 10.2196/jmir.1669] [Medline: 21955510]
12. The Lancet Gastroenterology & Hepatology. The potential of telemedicine in digestive diseases. The Lancet Gastroenterology & Hepatology 2019 Mar;4(3):185. [doi: 10.1016/s2468-1253(18)30359-5]

13. Heaney D, Caldow J, McClusky C, King G, Webster K, Mair F, et al. The introduction of a new consulting technology into the National Health Service (NHS) for Scotland. Telemed J E Health 2009 Jul;15(6):546-551. [doi: 10.1089/tmj.2009.0017] [Medline: 19519276]
14. Cooke M, Higgins J, Kidd P. Use of emergency observationassessment wards: a systematic literature review. Emerg Med J 2003;20(2):a. [doi: 10.1136/emj.20.2.138]
15. Lateef F, Anantharaman V. The short-stay emergency observation ward is here to stay. In: Am J Emerg Med. 2000;18(5); 2000:A.
16. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. Ann Emerg Med 2008 Aug;52(2):126-136. [doi: 10.1016/j.annemergmed.2008.03.014] [Medline: 18433933]
17. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One 2018 Jul 20;13(7):e0201016 [FREE Full text] [doi: 10.1371/journal.pone.0201016] [Medline: 30028888]
18. Ting DS, Gunasekeran DV, Wickham L, Wong TY. Next generation telemedicine platforms to screen and triage. Br J Ophthalmol 2020 Mar 03;104(3):299-300. [doi: 10.1136/bjophthalmol-2019-315066] [Medline: 31796427]
19. Hildebrandt DE, Westfall JM, Smith PC. After-hours telephone triage affects patient safety. J Fam Pract 2003 Mar;52(3):222-227. [Medline: 12620177]
20. Gunasekeran DV. Technology and chronic disease management. The Lancet Diabetes & Endocrinology 2018 Feb;6(2):91. [doi: 10.1016/s2213-8587(17)30441-2]
21. Ellis DG, Mayrose J, Phelan M. Consultation times in emergency telemedicine using realtime videoconferencing. J Telemed Telecare 2006 Jun 23;12(6):303-305. [doi: 10.1258/135763306778558187] [Medline: 17022838]
22. Brennan JA, Kealy JA, Gerardi LH, Shih R, Allegra J, Sannipoli L, et al. A randomized controlled trial of telemedicine in an emergency department. J Telemed Telecare 1998 Dec 02;4 Suppl 1(1_suppl):18-20. [doi: 10.1258/1357633981931911] [Medline: 9640721]
23. Kim SW, Madan J, Dritsaki M, Bryce C, Forjaz V, Fraser J, et al. Benefits and costs of digital consulting in clinics serving young people with long-term conditions: mixed-methods approach. JMIR Med Inform 2018 Oct 30;6(4):e48 [FREE Full text] [doi: 10.2196/medinform.9577] [Medline: 30377145]
24. van der Kop ML, Muhula S, Nagide PI, Thabane L, Gelmon L, Awiti PO, et al. Effect of an interactive text-messaging service on patient retention during the first year of HIV care in Kenya (WelTel Retain): an open-label, randomised parallel-group study. The Lancet Public Health 2018 Mar;3(3):e143-e152. [doi: 10.1016/s2468-2667(17)30239-6]
25. Gunasekeran DV, Wong TY. Artificial iintelligence in ophthalmology in 2020: a technology on the cusp for translation and implementation. Asia Pac J Ophthalmol (Phila) 2020;9(2):61-66. [doi: 10.1097/01.APO.0000656984.56467.2c] [Medline: 32349112]
26. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
27. Xie Y, Gunasekeran DV, Balaskas K, Keane PA, Sim DA, Bachmann LM, et al. Health economic and safety considerations for artificial intelligence applications in diabetic retinopathy screening. Transl Vis Sci Technol 2020 Apr;9(2):22.
28. Gunasekeran DV, Agrawal R, Testi I, Agarwal A, Mahajan S, Nguyen QD, Collaborative Ocular Tuberculosis Study (COTS) Group. Lessons in digital epidemiology from COTS-1: coordinating multicentre research across 10 countries using operational and technology innovation to overcome funding deficiencies. Ocul Immunol Inflamm 2020 Apr 22:1-7. [doi: 10.1080/09273948.2020.1744669] [Medline: 32320326]

Abbreviations

COVID-19: coronavirus disease 2019
SAVED: Safety and efficacy of follow-up for patients with abdominal pain using video consultation

Edited by G Eysenbach; submitted 11.12.19; peer-reviewed by P Więch, J Barros; comments to author 11.02.20; revised version received 20.02.20; accepted 29.02.20; published 15.06.20.

Please cite as:
Gunasekeran DV, Liu Z, Tan WJ, Koh J, Cheong CP, Tan LH, Lau CS, Phuah GK, Manuel NDA, Chia CC, Seng GS, Tong N, Huin MH, Dulce SV, Yap S, Ponampalam K, Ying H, Ong MEH, Ponampalam R
Evaluating Safety and Efficacy of Follow-up for Patients With Abdominal Pain Using Video Consultation (SAVED Study): Randomized Controlled Trial
J Med Internet Res 2020;22(6):e17417
URL: https://www.jmir.org/2020/6/e17417
doi:10.2196/17417
PMID:32459637

Original Paper

Communication Strategies Used to Obtain Clinical Histories Before Remotely Prescribing Antibiotics for Postal Treatment of Uncomplicated Genital Chlamydia: Service Evaluation

Hannah McCulloch[1], MSc; Jonathan Syred[1], MSc; Gillian Holdsworth[2], MSc, MPH, MBChB, FFPH; Chris Howroyd[2], BSc, MPhil; Elena Ardines[2], BSc; Paula Baraitser[1,2,3], BSc, MSc, MBBS, MD, MA, FFPH, FFSRH

[1]School of Population Health & Environmental Sciences, King's College London, London, United Kingdom
[2]SH:24, London, United Kingdom
[3]Department of Sexual Health and HIV, King's College Hospital, London, United Kingdom

Corresponding Author:
Paula Baraitser, BSc, MSc, MBBS, MD, MA, FFPH, FFSRH
School of Population Health & Environmental Sciences
King's College London
Weston Education Centre, 10 Cutcombe Rd
Denmark Hill Campus, SE5 9RJ
London
United Kingdom
Phone: 44 07525630865
Email: paula_baraitser@mac.com

Abstract

Background: Web-based services for testing of sexually transmitted infections are widely available across the United Kingdom. Remote prescriptions with medications posted home may support prompt treatment; however, the absence of face-to-face contact with clinicians raises clinical safety issues as medical history may not be accurately provided.

Objective: This service evaluation aimed to capture the use and explore the safety of 3 remote communication strategies employed within a web-based service offering remote prescriptions of antibiotics, delivered via post, for uncomplicated genital *Chlamydia trachomatis*. User acceptability and time-from-diagnosis-to-treatment were also obtained.

Methods: Three iterations of the service were compared, where medical history was collected via SMS text message, telephone, or a secure web form before a prescription was issued. We contacted users after they were issued a prescription and completed the medical history a second time via telephone, asking when they took their medication and how they felt about the service. The primary safety measure was agreement in information supplied at 2 assessments (ie, clinical and evaluation assessment) on key elements of safe prescribing: allergies, current medications, or contraindicating clinical conditions or symptoms. Agreement in information between clinical and evaluation assessment was summarized as a binary variable. Factors associated with the assessment agreement variable were explored using univariate and multivariate analysis. The secondary evaluation measures were recall of and adherence to instructions for taking medication, time-from-diagnosis-to-treatment, and acceptability of the web-based service.

Results: All web-based service users, resident in the London Boroughs of Lambeth and Southwark with a positive chlamydia diagnosis, who were eligible for and chose postal treatment between February 15, 2017, and October 24, 2017, were invited to participate in this service evaluation. Of 321 eligible users, 62.0% (199) participated. A total of 27.6% (55/199) users completed the clinical assessment via SMS text message, 40.7% (81/199) users via telephone, and 31.7% (63/199) users via a secure web form. Those who were assessed for prescription via SMS text message were less likely to have an agreement in safe prescribing information than those assessed via telephone (adjusted odds ratio [aOR] 0.22, 95% CI 0.08-0.61; *P*=.004). We found no statistically significant difference in odds of agreement between the web form and telephone assessment (aOR 0.50, 95% CI 0.17-1.43; *P*=.20). Median time-to-treatment was 4 days (IQR 3-5.5). In addition, 99.0% (196/199) of users reported understanding remote communication, and 89.9% (178/198) would use the service again.

Conclusions: Postal treatment is an acceptable and rapid treatment option for uncomplicated genital chlamydia. Clinical assessment via SMS text message before remote prescription may not be accurate or sufficient. As health care is delivered via the web, strategies that support safe remote prescribing are increasingly important, as is their evaluation, which should be robust and carefully considered.

(J Med Internet Res 2020;22(6):e15970) doi:10.2196/15970

KEYWORDS

sexual health; electronic prescribing; telemedicine; remote consultation; sexually transmitted diseases, bacterial

Introduction

Background

Web-based services for testing for sexually transmitted infections (STI) with test kits sent home are widely implemented in the United Kingdom [1,2]. In these services, users order a test kit from a web-based service, take the samples themselves, and post them to a laboratory with results given via SMS text message or telephone. There is evidence that they increase testing uptake in comparison with clinic-based services [3], but the clinical and public health impact of better access to testing requires prompt treatment of the diagnosed infections to prevent onward transmission. Remote prescriptions with medication posted home may facilitate prompt treatment.

Remote prescriptions without face-to-face health professional contact raise safety concerns [4]. The UK General Medical Council advises that clinicians should be satisfied in their ability to adequately assess the patient’s condition through satisfactory medical and drug history, including allergies before remote prescribing, and recommends consideration of the limitations of the medium of communication used to ascertain this information [5]. Health Care Improvement Scotland and the UK National Prescribing Centre have made similar recommendations [4,6,7], advocating a cautious approach to remote prescribing. This is particularly important when prescribing treatment for chlamydia within a web-based sexual health service because (1) there is no face-to-face contact throughout the process of testing, diagnosis, and treatment; (2) sexual health service records are not linked to the primary care record making it difficult to cross-check information, and (3) instructions for use and partner notification are provided remotely.

In this context, it is therefore, particularly important that there is effective remote communication to obtain accurate and sufficient medical history for prescribing, and to ensure correct use of medication and partner notification.

Research has documented poor practice in remote prescribing with medications available without a prescription or an appropriate medical history [8,9]; there are obvious health risks associated with this poor practice [10,11]. However, remote prescribing is also delivered by highly regulated providers where doctors obtain clinical histories through telephone, video, or email consultations [12,13]. There is good evidence that self-completed digital forms collect reliable data on past medical histories for diagnosis [14-17], but little evidence on their value before prescribing. As many health systems move toward a *digital first* approach (see, eg, the National Health Service, NHS long-term plan [18]) and increasingly use digital or web-based forms for clinical histories, we predict a need to understand their ability to generate accurate information for prescribing.

We developed and piloted remote prescribing and postal treatment of uncomplicated chlamydia infection within a web-based sexual health service. The process and outcomes of testing have been described elsewhere [3,19]. This service was evaluated during its implementation and development to monitor its clinical safety. As the service evolved, 3 remote communication strategies were used to take clinical histories before the prescribing and postal delivery of treatment of uncomplicated chlamydia. As the service evolved, the accuracy of the information obtained through each remote communication strategy was checked by telephoning users after they had received their prescriptions to recheck the information that they had provided. We sought to understand the safety issues highlighted by this service improvement activity. To do this, we described 3 cycles of service development to understand:

- How different media of assessment support accurate medical histories for prescribing
- To what extent service users understood remotely delivered instructions for use
- How different strategies for remote prescription and treatment impacted on time-to-treatment

The Service Evaluated

Sexual Health 24 hours a day (SH:24) is an NHS commissioned web-based service that specializes in system transformation in sexual health services through agile and design-led thinking, using an iterative process of build/test/learn for innovation. SH:24 offers postal self-sampling test kits for chlamydia, gonorrhea, HIV, and syphilis. Test kits are ordered through a website, with those reporting symptoms of infection referred to local clinics. Test kits are sent by post in discreet packaging that includes the materials for self-sampling—urine for men and vaginal samples for women—and a finger-prick blood sample for HIV and syphilis. The user posts the samples to a laboratory for testing. Chlamydia, gonorrhea, syphilis, and negative HIV test results are delivered via SMS text message. Reactive HIV test results are delivered via telephone with referral to local clinics for confirmatory testing.

Those with positive chlamydia results who are older than 18 years and asymptomatic are offered postal chlamydia treatment and complete a remote assessment to check their eligibility for the appropriate antibiotic. If eligible, an electronic prescription is issued by a clinician, and the medication is dispensed, packaged, and dispatched by a regulated pharmacy through the Royal Mail registered (but not signed for) postal service. If ineligible for the treatment offered by the web-based service, which at the time of evaluation was a single 1 g dose of

Azithromycin (SDA) users would be directed to local clinic services for treatment.

During and in response to the evaluation, the media of communication for assessment before the remote prescription changed 3 times.

1. During the first iteration of the service, eligibility was assessed via SMS text message. A single message rather than multiple messages was used as a break in communication during a process (eg, a loss of phone connection) would disrupt the assessment process. The message stated the following:

> *Your chlamydia result is positive. We would like to post you a single dose treatment. You would receive it the next working day. First, we need to know: Are you taking medicines or allergic to any medicines, soya or peanuts? (this is important because the medicine may trigger your allergy or interact with other medicines). Do you have liver, kidney, heart problems, or myasthenia gravis? (this is important because treatment may worsen your condition). Are you pregnant or breastfeeding? (this is important as the treatment may affect your baby). Do you have any symptoms: fever, joint pain, pelvic (lower abdominal) pain, or anal pain? (this is important as you may need a different treatment). If you answered no to ALL of these questions, reply NO. If you answered yes to ANY of these questions, reply YES. If you would like to go to a clinic for treatment instead, please reply CLINIC. Text back if you would like help. Thanks, SH:24.*

Those that responded no were then sent this message.

> *Thank you. Your treatment is now being prepared. You have told us that you are not taking any medicines, you do not have any medical problems (liver, kidney, heart, myasthenia gravis) and you do not have any symptoms of chlamydia. Text back if you have any queries or questions we always prefer to answer queries to prevent any future problems. Thanks, SH:24*

2. During the second iteration, the same questions were asked by a clinician over the telephone with responses recorded in the web-based clinical record.

3. During the third iteration, the same questions were asked via a secure web form, which was self-completed by service users.

Instructions for medication use, including abstinence from sexual intercourse for 7 days after commencing treatment and partner notification was provided in written form with the medication. The change from iteration 1 to iteration 2 occurred because a user reported an allergy in the evaluation that was not reported in the clinical assessment. The change from iteration 2 to iteration 3 occurred because it was felt to be time inefficient for the clinical team. Iteration 3 is the system currently implemented.

Methods

Design and Data Collection

All SH:24 users, resident in the London Boroughs of Lambeth and Southwark, with a positive chlamydia diagnosis who were eligible for and chose postal treatment between February 15, 2017, and October 24, 2017, were invited to take part in this service evaluation. Data were collected through a standardized telephone questionnaire delivered by a trained researcher after the medication had been prescribed and dispatched. This questionnaire documented relevant medical history, medication history (including allergies), time-to-treatment, and user experience. The full questionnaire can be found in Multimedia Appendix 1.

Each user, in effect, received the medical eligibility questions twice—once during their clinical assessment and once during the service evaluation. In iterations 1 and 3, there was a difference in the mode of communication between the clinical assessment and the evaluation assessment. During iteration 2, the mode of questioning before and after treatment delivery was the same. This is illustrated in Figure 1. As this was a service evaluation, there were no power calculations to determine the number of people in each arm. The service developed in response to real-time analysis of the evaluation rather than according to a research plan.

Measures

We refer to the preprescription assessment as the clinical assessment and the postprescription assessment as the evaluation assessment. The primary safety measure was agreement in information supplied at the 2 assessments on key elements of safe prescribing: allergies, current medications, or contraindicating clinical conditions or symptoms. The secondary evaluation measures were recall of and adherence to instructions for taking the medication, time-from-diagnosis-to-treatment, and acceptability of the postal treatment service.

Analysis

Descriptive statistics were used to summarize participant characteristics using parametric and, where appropriate, nonparametric tests. Agreement in key safe prescribing information supplied at the clinical assessment and the evaluation assessment was summarized as 1 binary variable, where matching information between assessments was given a value of 1, and conflicting or disparate information was given a value of 0. Logistic regression was then used to determine factors associated with agreement in key safe prescribing information. Univariate analysis was used to determine its relationship with demographic characteristics, previous clinic attendance, and method of clinical assessment, and variables significant at a 0.05 level were compiled to create a multivariable logistic model. A Kruskal-Wallis H test was conducted to determine the relationship between median time-to-treatment and method of clinical assessment.

Figure 1. Data collection points and methods.

Clinical assessment		*Evaluation assessment*
Iteration 1 — via SMS	Postal treatment	Assessment via telephone
Iteration 2 — via telephone		Assessment via telephone
Iteration 3 — via online form		Assessment via telephone

Ethical Considerations

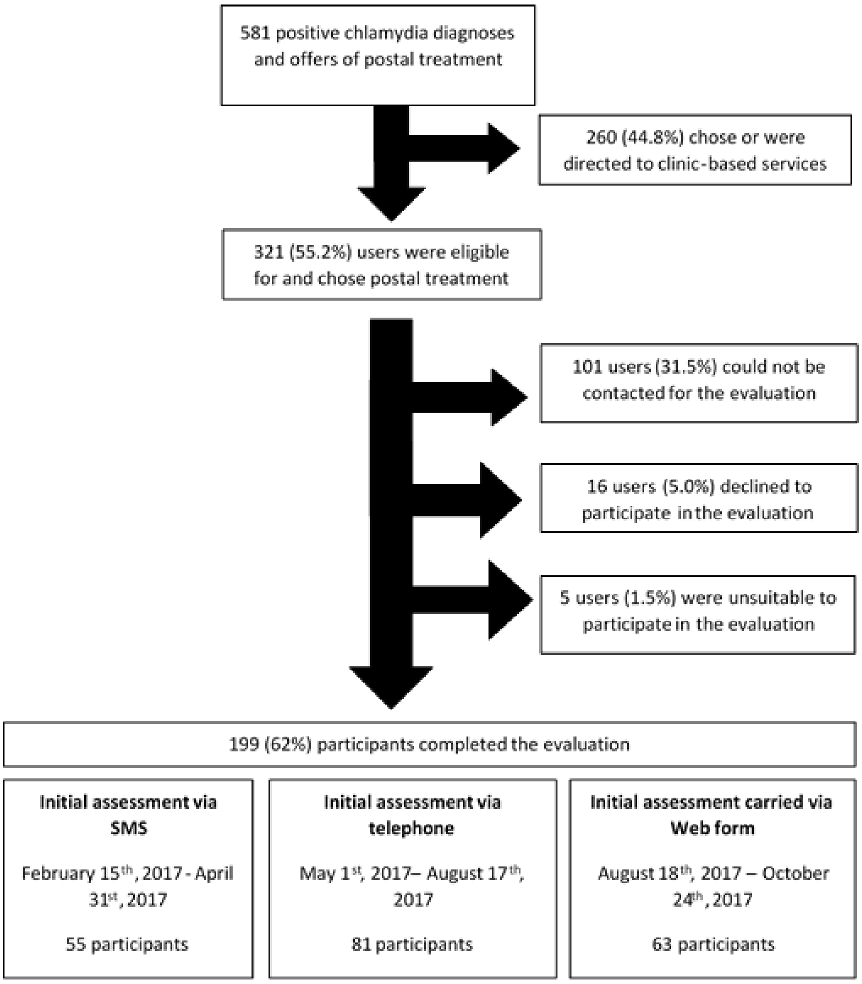
This work observed a build-test-learn cycle based on a human-centred design approach to service development carried out by a digital sexual health service, which adapted strategies that are already used by UK-registered web-based pharmacies. As a service evaluation, it focuses on auditable outcomes such as compliance with medical advice and processes, as well as time to treatment. For this service evaluation project, we did not request a formal ethics review; however, we conducted the evaluation in line with standard ethical procedures. To maintain confidentiality, users were asked to confirm the first line of their address before any information was disclosed about the nature of the telephone call. They were given full information about the evaluation and given the option to participate. Paper surveys, filled in by the researcher while on the phone, were marked with the user's unique identifier and stored in a locked filing cabinet. Unique identifiers could only be linked to identifiable data by accessing the web-based service.

Results

Evaluation Uptake

During the study period, there were 581 chlamydia diagnoses and offers of postal treatment, of which 260 (44.8%) chose to access or were directed to clinic-based services (mainly because they had symptoms). Of the 321 who chose postal treatment, 199 (62.0%) participated in the evaluation (101 could not be contacted, 16 declined to participate, and 5 were excluded—3 repeat users, 1 could not confirm their address, and 1 had decided to attend a clinic). Of the 199 participants, 55 (27.6%) completed the clinical assessment via SMS text message, 81 (40.7%) via telephone, and 63 (31.7%) via a secure web form (Figure 2). The median time between the clinical assessment and the evaluation assessment was 17 days (IQR 10-35).

Figure 2. Evaluation flowchart.



Participant Characteristics

The sociodemographic and clinic attendance characteristics of those who did and did not take part in the evaluation are described in Table 1.

Table 1. Sociodemographic characteristics of individuals who ordered postal treatment for chlamydia (N=321).

Sociodemographic characteristic	Sample (n=199), n (%)	Population (N=321), n (%)
Age group (years)		
18-19	21 (10.6)	37 (11.5)
20-24	83 (41.7)	126 (39.3)
25-29	54 (27.1)	95 (29.6)
30-34	26 (13.1)	39 (12.1)
35+	15 (7.5)	24 (7.5)
Gender		
Male	99 (49.7)	168 (52.3)
Female	100 (50.3)	153 (47.7)
Ethnicity		
White	76 (38.2)	143 (44.6)
Black British or Black Caribbean or Black African	73 (36.7)	103 (32.1)
Asian or Asian British	8 (4.0)	12 (3.7)
Mixed or multiple groups	24 (12.1)	38 (11.8)
Other or prefer not to say	18 (9.0)	25 (7.8)
Sexual orientation		
Heterosexual	176 (88.5)	288 (89.7)
Homosexual	5 (2.5)	7 (2.2)
Bisexual	14 (7.0)	17 (5.3)
Prefer not to say	4 (2.0)	9. (2.8)
Index of multiple deprivation quintile		
1 (most deprived)	66 (33.2)	108 (33.6)
2	83 (41.7)	127 (39.6)
3	32 (16.1)	56 (17.5)
4	13 (6.5)	19 (5.9)
5 (least deprived)	5 (2.5)	11 (3.4)
Ever attended a sexual health clinic?		
Yes	147 (73.9)	227 (70.7)
No	52 (26.1)	94 (29.3)
Attended a sexual health clinic in the last 12 months?[a]		
Yes	77 (38.7)	109 (34.0)
No	94 (47.2)	150 (46.7)
Did not answer	28 (14.1)	62 (19.3)

[a] Was not answered by all participants.

Agreement in Key Safe Prescribing Information

We considered all discrepancies between the clinical and evaluation assessment. New clinical information reported at the evaluation assessment but not the clinical assessment was reported to the web-based service clinical team who managed it according to the clinical incident policy.

During the clinical assessment, 154 participants reported no contraindications, and all except 2 reported the same information at the evaluation assessment. The 2 exceptions were 2 women who were taking the contraceptive pill.

During the evaluation assessment, 45 individuals reported potential contraindications, and 31 of these omitted this from their clinical assessment (14 unreported medications, 6 allergies, 1 previous operation, and 10 genital symptoms). One of these contraindications, an allergy to peanuts, could have caused a serious interaction with the medication prescribed.

Of the 33 discrepancies between clinical and evaluation assessments, 15 (46%) occurred when the clinical assessment was completed via SMS text message, 8 (24%) via telephone, and 10 (30%) via the secure web form. Univariate analysis showed that method of clinical assessment, gender, and sexuality were independently associated with agreement between the clinical and evaluation assessment. When combined in a multivariable model, method of initial assessment, gender, and sexuality remained statistically significantly associated with agreement between the 2 assessments.

The adjusted odds of agreement between assessments were 78% less among participants who completed their clinical assessment via SMS text messaging compared with those who completed this via telephone. This result was statistically significant (see Table 2). We found no statistically significant difference between the odds of agreement between assessments completed via telephone and via the secure web form.

Of those who took their medication, 91.1% (175/192) correctly recalled advice to abstain from sex for 7 days after they and their partner had commenced treatment, 93.4% (184/197) recalled advice on partner notification, and 89.3% (176/197) had notified partners.

Table 2. Crude and adjusted odd ratios of agreement in reporting information.

Exposure variable	Crude odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)[a]	*P* value
Method of initial assessment				
Call	1 (reference[b])	—[b]	1 (reference)	—[b]
SMS text message	0.29 (0.11-0.75)	.01	0.22 (0.08-0.61)	.004
Secure web form	0.58 (0.21-1.57)	.28	0.50 (0.17-1.43)	.20
Ever attended a clinic?				
Yes	1 (reference)	—[b]	1 (reference)	—[b]
No	2.21 (0.81-6.07)	.12	—[d]	—[d]
Gender				
Female	1 (reference)	—[b]	1 (reference)	—[b]
Male	2.66 (1.19-5.93)	.02	2.75 (1.16-6.56)	.02
Age (years)				
18-19	1 (reference)	—[b]	1 (reference)	—[b]
20-24	1.68 (0.52-5.40)	.38	—[d]	—[d]
25-29	2.10 (0.58-7.55)	.26	—[d]	—[d]
30-34	1.31 (0.32-5.32)	.70	—[d]	—[d]
35+	1.25 (0.25-6.29)	.78	—[d]	—[d]
Ethnicity				
White	1 (reference)	—[b]	1 (reference)	—[b]
Black British/Caribbean/African	0.87 (0.37-2.05)	.74	—[d]	—[d]
Asian/Asian British	1.31 (0.15-11.66)	.81	—[d]	—[d]
Mixed/multiple groups	0.56 (0.19-1.71)	.31	—[d]	—[d]
Other/prefer not to say	3.19 (0.39-26.26)	.28	—[d]	—[d]
Sexuality				
Heterosexual	1 (reference)	—[b]	1 (reference)	—[b]
Homosexual	0.69 (0.07-6.45)	.75	0.34 (0.03-3.65)	.37
Bisexual	0.23 (0.07-0.72)	.01	0.19 (0.06-0.66)	.009
Prefer not to say	—[c]	—[c]	—[c]	—[c]
Index of multiple deprivation q uintile				
1 (most deprived)	1 (reference)	—[b]	1 (reference)	—[b]
2	1.10 (0.47-2.56)	.83	—[d]	—[d]
3	3.33 (0.70-15.90)	.13	—[d]	—[d]
4	0.36 (0.10-1.28)	.13	—[d]	—[d]
5 (least deprived)	—[c]	—[c]	—[d]	—[d]

[a] Only variables where crude odds ratios were found to be significant ($P<.05$) were carried into the multivariable model.

[b] Reference group.

[c] Not applicable, as variable was not carried into the multivariable model.

[d] Not applicable, as no sample participants within these categories.

Time-From-Diagnosis-to-Treatment

From the 192 participants who took their treatment, median time-to-treatment was 4 days (IQR 3-5.5). A total of 91.7% (176/192) of those evaluated were treated within 7 days of receiving their positive diagnosis.

A Kruskal-Wallis H test was used to analyze the relationship between time-to-treatment and the medium of the clinical assessment as this variable was not normally distributed. Of the 192, 55 were assessed via SMS text message, 76 via telephone, and 61 via web form. There was a statistically significant difference between time-to-treatment by method of clinical assessment (H_2=24.169; P<.001), with a median time-to-treatment of 2.5 days for SMS text messaging, 4 days for web form, and 4.5 days for telephone call. Similarly, statistically significant differences were found between SMS text message and telephone call (H_2=21.990; P<.001) and SMS text message and web form (H_2=14.702; P<.001). There was no statistically significant difference in the median time-to-treatment between web form and telephone call.

User Acceptability

The majority of users (185/197, 93.9%) were happy with the information provided on treatment. Almost all users reported that they could both understand (196/199, 99.0%) and were comfortable (196/198, 98.5.0%) with the SMS text messages they received from the service. 19.4% (38/196) contacted SH:24 for support, and 89.9% (178/198) said they would use the web-based service again.

Discussion

Principal Findings

This service evaluation documented an iterative process of testing 3 methods of clinical assessment before remote prescribing. It showed that users are significantly more likely to report accurate clinical histories during a telephone consultation than through a single SMS text message. We did not find a statistically significant difference between the accuracy of clinical histories reported via telephone and via the secure web form. It showed that treatment via a web-based sexual health service in general, and SMS text message assessment in particular, supports rapid treatment that is well within the standards set in current guidance; the National Chlamydia Screening Programme (NCSP) audit standard aims for 95% of patients to have a time-to-treatment of 6 weeks or less [20,21]. An NCSP audit of chlamydia treatment services carried out in 2017 [22] found that 92.0% of the sample met this standard, whereas within this service evaluation, 91.7% (176/192) of those evaluated were treated within 7 days of receiving their positive diagnosis. This is comparable with findings from Estcourt et al’s [23] exploratory studies into an electronic sexual health clinic system for management, prevention, and control of STIs, where results were delivered via a secure web portal and treatment was collected from a local pharmacy.

Our findings are important as treatment is increasingly part of routine web-based sexual health services. This is reflected both in the use of the web-based service evaluated here and quality standards and guidelines recently released by relevant professional bodies.

Since the start of the service in 2017, 5130 chlamydia treatments have been delivered via SH:24 by post in 14 different areas of the country. Postal treatment is highly acceptable to service users, with 80% of those offered treatment by post taking up this offer. More information about the uptake and number of orders can be found in Multimedia Appendix 2. The service now uses the secure web form to assess eligibility, and SMS text messaging or telephone are no longer used, except to gain more detail on contraindications that have been reported in the form.

Since the development of the service, the Faculty of Sexual and Reproductive Healthcare and the British Association for Sexual Health and HIV (BASHH) have published the first quality standards for online and remote providers of sexual and reproductive health care [24]. These standards not only reiterate the General Medical Council’s (GMC’s) recommendations on safety and remote prescribing [5] but also highlight the *safety nets* services should have to navigate and mitigate risks that may come with remote consultations [24]. The service evaluated here has developed a *2-way process* around the secure web form that is now used as the clinical assessment for eligibility in the web-based remote chlamydia treatment service where free textboxes within the web form allow users to share, for example, any medication they may be taking. Furthermore, as per the GMC guidelines, they now ask patients for their consent to share treatment information with their general practitioner [25].

Other changes in the service since this evaluation reflect updated clinical guidelines; during the evaluation users were offered SDA as first-line treatment; however, in response to BASHH recommendations [21], the first-line treatment offer for uncomplicated urogenital, pharyngeal, and rectal chlamydia infections is now Doxycycline 100 mg bd for 7 days.

Interpretation

Higher risks associated with the SMS text message assessment provide important learning. New technologies such as cross-platform messaging services (eg, WhatsApp) are increasingly used by businesses for secure communication with their customers and strategies for assessment before medical prescribing may develop in this direction. We recommend further investigation into the risks and benefits of these approaches. The SMS text messaging strategy used here was limited to a single (rather long) message that included multiple questions and required a single answer. This is suboptimal and is a function of the limitations of SMS text messaging technology. It was not possible to break up the message into individual SMS text messages in case the connection was lost halfway through the interaction. SMS text messages remain undelivered when the connection breaks, unlike, for example, WhatsApp, where messages are delivered as soon as the connection is resumed. web forms are commonly used for history taking before prescriptions. This evaluation suggests that this is a promising approach, but our sample size was insufficient to provide significant results, and future research should explore their performance in comparison with telephone and face-to-face assessment.

Our analysis assumes that telephone assessment by an experienced clinician is safe as this is standard practice, but it also suggests that even this method is associated with inaccuracies. Even face-to-face consultations are associated with medication errors (usually prescribing errors) in hospitals varying between 2% and 14% [26]. It is likely that no method is consistently reliable in obtaining an accurate history. Situations where the clinical history can be checked against previous health records have obvious advantages, but this is not possible in sexual health where information traditionally is not linked to the rest of the medical record as a strategy to maintain confidentiality [27].

Limitations

A small evaluation of a rapidly evolving service, this service evaluation did not calculate sample size when conceived. Participants were not allocated to a method of clinical assessment; designed to judge current care provided, the service evaluation followed an evolving service, where 3 iterations of the service design allowed for retrospective exploration of factors associated with agreement in safe prescribing information, including method of clinical assessment. Although, even with a small sample, we did find a difference in odds of agreement between SMS text messaging and telephone call, a robust, considered piece of research, with a calculated sample size, would be needed to fully explore the relationship between method of clinical assessment and agreement in safe prescribing information. As methods were tested consecutively, users were not randomly allocated to a particular method. Hence, those carrying out the follow-up assessment may not have been blinded; a possible source of bias.

We recognize the role that reporting bias may have played in user evaluation assessments. Users asked to reflect on their symptoms, once they have been given a diagnosis, may recognize symptoms they previously may have been deemed unrelated. Several women reported abdominal pain in the evaluation assessment, that they had considered period pain before they received their positive diagnosis. Preexisting literature highlights poor knowledge of nature and symptoms of chlamydia infection [28-30]. In a study of 18- to 24-year olds, a third of respondents were unaware of the asymptomatic nature of chlamydia infection, and 80.2% of respondents failed to recognize lower abdominal pain as a potential symptom in women [30].

Although most participants of this service evaluation stated a preference for a remote treatment service, as the sample consists solely of those who self-identify as asymptomatic, this should not be generalized to everyone seeking treatment. Sexual health clinics remain an integral element of STI treatment and care, and web-based services should be well integrated with face-to-face services [31,32]. However, considering that at least 70% of women and 50% of men infected with chlamydia are asymptomatic [33,34], this shows acceptability in a large proportion of those who need treatment.

Conclusions and Recommendations

Postal treatment is an acceptable and rapid treatment option for uncomplicated genital chlamydia. Clinical assessment via SMS text message before remote prescription may not be accurate or sufficient. As health care is delivered, strategies that support safe remote prescribing are increasingly important, as is their evaluation, which should be robust and carefully considered.

Acknowledgments

This service evaluation was undertaken as part of a multidisciplinary evaluation of web-based sexual health services funded by Guys and St Thomas' Charity.

Authors' Contributions

PB, JS, GH, and CH contributed to the study conception. AS, JS, PW, and HM were responsible for data collection. HM contributed to the analysis. HM and PB contributed to the write-up.

Conflicts of Interest

PB, GH, and CH are directors of SH:24 (a not-for-profit community interest company). EA is also employed by SH:24. Since October 2019, HM has been working as a research associate at SH:24.

Multimedia Appendix 1
Service evaluation questions.
[DOCX File , 13 KB - jmir_v22i6e15970_app1.docx]

Multimedia Appendix 2
Information on chlamydia treatment service provision.
[DOCX File , 13 KB - jmir_v22i6e15970_app2.docx]

References

1. Cleary M, O'Sullivan J. P145 London sexual health transformation programme. Sex Transm Infect 2017;93:A64-A65 [FREE Full text] [doi: 10.26226/morressier.58ef94ddd462b80290b50420]
2. NHS England. 2019. Find Chlamydia - Free Online Tests for u-25s Services URL: https://www.nhs.uk/Service-Search/Chlamydia-free-online-tests-for-u-25s/LocationSearch/344 [accessed 2019-01-25]
3. Wilson E, Free C, Morris TP, Syred J, Ahamed I, Menon-Johansson AS, et al. Internet-accessed sexually transmitted infection (e-STI) testing and results service: A randomised, single-blind, controlled trial. PLoS Med 2017 Dec;14(12):e1002479 [FREE Full text] [doi: 10.1371/journal.pmed.1002479] [Medline: 29281628]
4. Healthcare Improvement Scotland. 2014 Apr 15. Implementing an Electronic Prescribing and Medicines Administration System: A Good Practice Guide URL: http://www.healthcareimprovementscotland.org/our_work/technologies_and_medicines/electronic_prescribing/good_practice_guide.aspx [accessed 2018-08-03]
5. General Medical Council. 2013. Good Practice in Prescribing and Managing Medicines and Devices URL: https://www.gmc-uk.org/Prescribing_guidance.pdf_59055247.pdf [accessed 2018-08-10]
6. Martin D, Millson D, Avery A, Smart S. The King's Fund. 2011. The Quality of GP Prescribing URL: https://www.kingsfund.org.uk/sites/files/kf/field/field_document/quality-gp-prescribing-gp-inquiry-research-paper-mar11.pdf [accessed 2019-01-26]
7. National Prescribing Centre. Pharmaceutical Services Negotiating Committee (PSNC). Liverpool: National Prescribing Centre; 2008 Sep. Dispensing with Repeats: A Practical Guide to Repeat Dispensing. Second Edition URL: http://archive.psnc.org.uk/data/files/repeat_dispensing/npcdispensing_with_repeats.pdf [accessed 2019-02-02]
8. Boyd S, Moore L, Gilchrist M, Costelloe C, Castro-Sánchez E, Franklin B, et al. Obtaining antibiotics online from within the UK: a cross-sectional study. J Antimicrob Chemother 2017 May 1;72(5):1521-1528 [FREE Full text] [doi: 10.1093/jac/dkx003] [Medline: 28333179]
9. Alwon B, Solomon G, Hussain F, Wright D. A detailed analysis of online pharmacy characteristics to inform safe usage by patients. Int J Clin Pharm 2015 Feb;37(1):148-158. [doi: 10.1007/s11096-014-0056-1] [Medline: 25564180]
10. General Pharmaceutical Council. 2018 Jun. Discussion Paper on Making Sure Patients and the Public Obtain Medicines and Other Pharmacy Services Safely Online URL: https://www.pharmacyregulation.org/sites/default/files/document/discussion_paper_on_safe_online_pharmacy_june_2018_0.pdf [accessed 2019-02-07]
11. Care Quality Commission. 2018. The State of Care in Independent Online Primary Health Services URL: https://www.cqc.org.uk/publications/major-report/state-care-independent-online-primary-health-services [accessed 2019-02-07]
12. Lloyds Pharmacy. NHS Experienced Clinicians | LloydsPharmacy. 2019. Chlamydia Treatment URL: https://onlinedoctor.lloydspharmacy.com/uk/sti-treatments/chlamydia-treatment [accessed 2019-03-04]
13. Superdrug. Superdrug Online Doctor. 2019. Chlamydia Treatment URL: https://onlinedoctor.superdrug.com/chlamydia-treatment.html [accessed 2019-01-25]
14. Tideman R, Chen M, Pitts M, Ginige S, Slaney M, Fairley C. A randomised controlled trial comparing computer-assisted with face-to-face sexual history taking in a clinical setting. Sex Transm Infect 2007 Feb;83(1):52-56 [FREE Full text] [doi: 10.1136/sti.2006.020776] [Medline: 17098771]
15. Kurth AE, Martin DP, Golden MR, Weiss NS, Heagerty PJ, Spielberg F, et al. A comparison between audio computer-assisted self-interviews and clinician interviews for obtaining the sexual history. Sex Transm Dis 2004 Dec;31(12):719-726. [doi: 10.1097/01.olq.0000145855.36181.13] [Medline: 15608586]
16. Richens J, Copas A, Sadiq ST, Kingori P, McCarthy O, Jones V, et al. A randomised controlled trial of computer-assisted interviewing in sexual health clinics. Sex Transm Infect 2010 Aug;86(4):310-314. [doi: 10.1136/sti.2010.043422] [Medline: 20551234]
17. Ghanem KG, Hutton HE, Zenilman JM, Zimba R, Erbelding EJ. Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. Sex Transm Infect 2005 Oct;81(5):421-425 [FREE Full text] [doi: 10.1136/sti.2004.013193] [Medline: 16199744]
18. NHS England. NHS Long term plan. 2019. URL: https://www.longtermplan.nhs.uk/ [accessed 2019-01-25]
19. Wilson E, Leyrat C, Baraitser P, Free C. Does internet-accessed STI (e-STI) testing increase testing uptake for chlamydia and other STIs among a young population who have never tested? Secondary analyses of data from a randomised controlled trial. Sex Transm Infect 2019 Dec;95(8):569-574 [FREE Full text] [doi: 10.1136/sextrans-2019-053992] [Medline: 31175210]
20. Public Health England. The Government of UK. 2018 Nov 28. National Chlamydia Screening Programme Standards. Seventh Edition URL: https://www.gov.uk/government/publications/ncsp-standards [accessed 2019-01-25]
21. British Association for Sexual Health and HIV (BASHH). 2019. Standards for the Management of Sexually Transmitted Infections (STIs) URL: https://www.bashh.org/about-bashh/publications/standards-for-the-management-of-stis/ [accessed 2019-05-01]
22. Public Health England. The Government of UK. 2018. NCSP 2017 Audit Report. Audit of Standards: Turnaround Times, Partner Notification and Re-Testing URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/713332/NCSP_2017_audit_report.pdf [accessed 2019-05-01]
23. Estcourt CS, Gibbs J, Sutcliffe LJ, Gkatzidou V, Tickle L, Hone K, et al. The eSexual Health Clinic system for management, prevention, and control of sexually transmitted infections: exploratory studies in people testing for Chlamydia trachomatis.

Lancet Public Health 2017 Apr;2(4):e182-e190 [FREE Full text] [doi: 10.1016/S2468-2667(17)30034-8] [Medline: 29253450]
24. Faculty of Sexual and Reproductive Healthcare (FSRH), British Association for Sexual Health and HIV (BASHH). The Faculty of Sexual and Reproductive Healthcare. London, UK: Joint BASHH/FSRH Standard; 2019 Jan. Standards for Online and Remote Providers of Sexual and Reproductive Health Services URL: https://www.fsrh.org/standards-and-guidance/documents/fsrhbashh-standards-for-online-and-remote-providers-of-sexual/ [accessed 2019-05-01]
25. General Medical Council. 2018. Remote Consultations URL: https://www.gmc-uk.org/ethical-guidance/ethical-hub/remote-consultations [accessed 2019-01-06]
26. Williams D. Medication errors. J R Coll Physicians Edinb 2007;37(4):343-347 [FREE Full text]
27. Department of Health and Social Care, Public Health England. The Government of UK. 2018. Integrated Sexual Health Services: A Suggested National Service Specification URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/731140/integrated-sexual-health-services-specification.pdf [accessed 2019-07-15]
28. Greaves A, Lonsdale S, Whinney S, Hood E, Mossop H, Olowokure B. University undergraduates' knowledge of chlamydia screening services and chlamydia infection following the introduction of a National Chlamydia Screening Programme. Eur J Contracept Reprod Health Care 2009 Feb;14(1):61-68. [doi: 10.1080/13625180802434884] [Medline: 19241303]
29. Chaudhary R, Heffernan CM, Illsley AL, Jarvie LK, Lattimer C, Nwuba AE, et al. Opportunistic screening for Chlamydia: a pilot study into male perspectives on provision of Chlamydia screening in a UK university. J Public Health (Oxf) 2008 Dec;30(4):466-471. [doi: 10.1093/pubmed/fdn060] [Medline: 18653445]
30. Lorimer K, Hart GJ. Knowledge of Chlamydia trachomatis among men and women approached to participate in community-based screening, Scotland, UK. BMC Public Health 2010 Dec 30;10:794 [FREE Full text] [doi: 10.1186/1471-2458-10-794] [Medline: 21192793]
31. National Voices. 2014. Supporting Self-Management: Summarising Evidence From Systematic Reviews URL: https://www.nationalvoices.org.uk/sites/default/files/public/publications/supporting_self-management.pdf [accessed 2019-07-15]
32. Turner KM, Zienkiewicz AK, Syred J, Looker KJ, de Sa J, Brady M, et al. Web-based activity within a sexual health economy: observational study. J Med Internet Res 2018 Mar 7;20(3):e74 [FREE Full text] [doi: 10.2196/jmir.8101] [Medline: 29514776]
33. National Institute for Health and Care Excellence. 2019. Chlamydia - Uncomplicated Genital URL: https://cks.nice.org.uk/chlamydia-uncomplicated-genital#!topicSummary [accessed 2019-07-15]
34. Scottish Intercollegiate Guidelines Network. Management of genital Chlamydia trachomatis infection: A national clinical guideline. 2009. NHS Quality Improvement Scotland URL: https://www.sign.ac.uk/assets/sign109.pdf [accessed 2020-06-01]

Abbreviations

BASHH: British Association for Sexual Health and HIV
GMC: General Medical Council
NCSP: National Chlamydia Screening Programme
NHS: National Health Service
SDA: single dose of Azithromycin
SH: 24: Sexual Health 24 hours a day
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 29.08.19; peer-reviewed by J Gibbs, SC Pan; comments to author 14.12.19; revised version received 16.01.20; accepted 24.02.20; published 17.06.20.

Please cite as:
McCulloch H, Syred J, Holdsworth G, Howroyd C, Ardines E, Baraitser P
Communication Strategies Used to Obtain Clinical Histories Before Remotely Prescribing Antibiotics for Postal Treatment of Uncomplicated Genital Chlamydia: Service Evaluation
J Med Internet Res 2020;22(6):e15970
URL: http://www.jmir.org/2020/6/e15970/
doi:10.2196/15970
PMID:32554379

Original Paper

Perspectives of Psychotherapists and Psychiatrists on Mental Health Care Integration Within Primary Care Via Video Consultations: Qualitative Preimplementation Study

Mariell Hoffmann[1], BA, MA; Michel Wensing[2], Prof Dr; Frank Peters-Klimm[2], Prof Dr; Joachim Szecsenyi[2], Prof Dr; Mechthild Hartmann[1], MA; Hans-Christoph Friederich[1], Prof Dr; Markus W Haun[1], BSc, MSc, MD

[1]Department of General Internal Medicine and Psychosomatics, Heidelberg University, Heidelberg, Germany
[2]Department of General Practice and Health Services Research, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Markus W Haun, BSc, MSc, MD
Department of General Internal Medicine and Psychosomatics
Heidelberg University
Im Neuenheimer Feld 410
Heidelberg, D-69120
Germany
Phone: 49 622156 ext 38396
Fax: 49 6221565988
Email: markus.haun@med.uni-heidelberg.de

Abstract

Background: Many patients with mental disorders remain untreated. Video-based mental health care demonstrates comparable effectiveness to face-to-face treatments and is a promising mode for delivering specialized care within primary care. Nevertheless, professionals struggle with implementing video consultations in their daily practice. Specifically, little is known about mental health specialists' acceptance of mental health video consultations in routine practice. The PROVIDE (ImPROving cross-sectoral collaboration between primary and psychosocial care: An implementation study on VIDEo consultations) project aims to improve cross-sectoral collaboration between primary and psychosocial care through implementing video consultations in primary care. To increase the uptake of video consultations, it is crucial to account for necessary prerequisites and to tailor interventions to the needs of the target group prior to implementation.

Objective: The aim of this study was to explore the acceptance of video consultations embedded in primary care from the perspectives of mental health specialists in Germany.

Methods: We conducted a qualitative, exploratory, preimplementation study in urban and rural counties. We conducted three semistructured focus groups with 11 mental health specialists. We used qualitative content analysis combining an inductive-deductive approach, applying the Tailored Implementation in Chronic Diseases (TICD) framework to the text material, which comprises individual health professional factors; patient factors; professional interactions; incentives and resources; capacity for organizational change; social, political, and legal factors; and guideline factors.

Results: Against the background of long waiting times and a shortage of mental health specialists, especially in rural areas, participants valued video consultations as a potential means to improve access to mental health care. With respect to the TICD framework domains, the participants most often discussed individual health professional factors, followed by patient factors. All participants highlighted the importance of a trusting relationship between the patient and the therapist and doubted whether such a relationship could be established through video consultations (11/11, 100%). However, participants considered mental health specialist video consultations to be particularly suited for patients in rural areas, those with impaired mobility, and those who may otherwise remain untreated (6/11, 55%). Most participants expected video consultations to help the aforementioned patient groups avoid tedious searching for an available therapist and save on travel time and, therefore, improve access to specialized care for patients (7/11, 64%). Moreover, the participants expected video consultations to improve collaboration with the family physician (6/11, 55%). Finally, participants identified organizational aspects, such as reliable scheduling, the duration of the individual consultation (9/11, 82%), and reimbursement conditions (7/11, 67%), as key drivers for the acceptance and adoption of the model.

Conclusions: While mental health specialists expect video consultations to improve access to specialized care for some patients, they consistently wonder whether such consultations can establish a trusting patient-therapist relationship. When implementing video consultations, these concerns should be addressed by training providers in managing technology-based treatment settings, with extra consideration for fostering the patients' and therapists' engagement.

Trial Registration: German Clinical Trials Register DRKS00012487; https://tinyurl.com/uhg2one

(J Med Internet Res 2020;22(6):e17569) doi:10.2196/17569

KEYWORDS

video consultations; videoconferencing; telehealth; integrated care; mental health; health services research; qualitative study; preimplementation

Introduction

Background

Many patients with common mental health conditions, such as depression, anxiety, or somatic symptom disorders, solely consult their family physicians. When seeking specialized mental health care, patients face the challenge of limited access, especially in rural areas. A significant number of these patients remain undiagnosed or receive psychiatric medication prescriptions, although most would prefer psychological treatment [1,2]. According to the American Psychiatric Association and the Academy of Psychosomatic Medicine, mental health care specialistss—namely, psychotherapists and psychiatrists—are required to provide patient-centered care and improve access of care (eg, by supporting integrated delivery and providing multiple points of access) [3]. Internet-based integrated care models such as video consultations are one way to address the demands for mental health care, provided that family physicians and mental health specialists can overcome system barriers and foster cross-sectoral communication. Several studies demonstrated acceptable feasibility and equal effectiveness of video consultations compared to face-to-face treatments in psychotherapy [4-10]. Additionally, video consultations can save costs and travel time [4,11]. Patients generally accept video consultations and are often satisfied with this mode of treatment delivery [6,8,12]. However, regarding the uptake of best evidence in practice, one consistent research finding is the gap between evidence-based interventions and their adoption in clinical practice. To bridge this gap, it is crucial to identify user acceptance of interventions and to explore prerequisites among the target group of intervention users to tailor the intervention to their needs and foster the adoption of best practice standards into routine care [13]. Several projects evaluated video-based integrated care models, but little is known regarding how mental health specialists, especially in Germany, perceive the acceptance of real-time video consultations between patients presenting to the primary care practice and off-site mental health specialists.

Mental Health Specialist Video Consultations Embedded in Primary Care

The PROVIDE (ImPROving cross-sectoral collaboration between primary and psychosocial care: An implementation study on VIDEo consultations) project aims to improve cross-sectoral collaboration between primary and psychosocial care through implementing video consultations in primary care. The intervention involves clinical diagnostics, care planning, and crisis management or brief psychotherapy and is limited to a maximum of five consultations. Specifically, mental health specialists will conduct video consultations with patients with depression and/or anxiety who are under the care of his or her primary care family physician. The PROVIDE project consists of three stages: first, we identified barriers and facilitators of stakeholders to determine their needs and tailor the model accordingly (preimplementation phase: PROVIDE-A). Supplementing our recently published work on the perceptions of family physicians regarding mental health specialist video consultations in primary care [14], this paper presents the results for psychotherapists and psychiatrists from the initial preimplementation phase of the PROVIDE project. Second, the video consultation model will be put into practice and evaluated regarding the acceptance and practicability within a pilot study (feasibility study: PROVIDE-B) [15]. Third, the model will be optimized where necessary and tested in a randomized controlled trial (RCT) with respect to the extent to which the burden of depression and anxiety of patients changes, as well as patient outcomes, as compared to patients who received usual care (implementation study, RCT: PROVIDE-C).

Purpose of the Study

This qualitative study from the preimplementation phase of PROVIDE-A aimed to (1) identify perceived acceptance of the intervention among mental health specialists and (2) shed light on prerequisites—namely, barriers and facilitators—for the implementation of mental health specialist video consultations.

Methods

Study Design

In order to assess the acceptance of mental health specialist video consultations, we conducted a cross-sectional, qualitative, exploratory preimplementation study with mental health specialists. The results of this study will be used to further tailor the intervention model to the practical needs of mental health specialists and will then be evaluated within a following feasibility study [15]. We conducted three focus groups with a total of 11 participants. By using focus groups, we were able to collect diverse perceptions and experiences, which focus group participants typically disclose to each other and on which they collectively reflect [16]. This study was approved by the Ethics Committee of Heidelberg Medical School (Reference S-197/2017) and was preregistered with the German Clinical Trials Register (DRKS00012487). We followed the COREQ

(COnsolidated criteria for REporting Qualitative research) guidelines for reporting study results (see Multimedia Appendix 1) [17]. The authors had full access to all the data in this study and take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Participants and Recruitment

We used purposeful sampling to obtain a broad range of perceptions [18]. To be eligible, mental health specialists had to be registered with the Association of Statutory Health Insurance Physicians in one of the five counties covered in this study; the five counties were selected from a total of 35 counties in Baden-Wuerttemberg, one of 16 German federal states. Overall, 583 mental health specialists were eligible. We chose a stratified randomized sample of two waves—first wave in June 2017 (N=100); second wave in September 2017 (N=195)—in order to recruit at least 10 participants for the focus groups. Due to the shortage of mental health specialists, particularly in rural areas, we aimed to explore potential distinctions between *rural and suburban* areas and *urban* areas. Therefore, we used the Degree of Urbanisation (DEGURBA) classification from the European Commission to distinguish the urban and rural counties [19]. Overall, 275 mental health specialists from cities, 261 mental health specialists from suburbs, and 47 mental health specialists from rural areas were eligible. We sent a written questionnaire to a total of 295 stratified randomized mental health specialists regarding their perception of mental health specialist video consultations (results not presented here) and invited them to participate in a focus group. If no response had been received within 3 weeks, we reminded all nonresponders via follow-up telephone calls (up to three calls). To schedule the focus groups, we contacted all 27 eligible mental health specialists, by telephone, who had expressed interest in participating. A total of 15 mental health specialists subsequently declined study participation, mainly due to time restrictions (9/15, 60%). One participant agreed to participate in a telephonic interview but is not included in the analysis, as we focus on collective reflections on the intervention. Overall, we conducted three semistructured focus groups with a total of 11 participants. We stopped data collection and started data analysis when no new insights occurred in the data and the content began to repeat. We offered a nonadvertised, individual monetary compensation of €50 (US $55) to all participants.

Focus Groups

For the development of a questionnaire guide (see Multimedia Appendix 2), we followed established recommendations [20] and used a team-based approach—M Hoffmann, sociologist; MW, implementation science; M Hartmann, psychotherapist; and MWH, internal medicine specialist—advancing to specific open-ended, jargon-free questions from the initial formulation of key questions in a logical sequence. The interviews focused on (1) how mental health care providers perceive current mental health care, (2) the acceptance of and intent to adopt mental health specialist video consultations, and (3) factors that promote or inhibit the use of the model. Before conducting the focus groups, we piloted the guide on one mental health specialist and modified it accordingly. The question guide was also reviewed after the first focus group with mental health specialists. The first author (M Hoffmann: PhD student, master's degree in sociology, and expertise in qualitative research) and the last author (MWH: MD, internal medicine specialist, senior researcher, and content expert for mental health services), who had no contact or relationship with any participant prior to the study, conducted the focus groups. The focus groups, which ranged from 2 to 5 participants and lasted 80-110 minutes, were conducted at Heidelberg University Hospital between August and November 2017; participants were from one urban and four rural or suburban counties. We used the question guide to encourage discussion and to collect comparable data through introducing similar discussion topics. To stimulate the discussion, the moderators presented a video clip illustrating the model using an example of a fictional patient suffering from depression. We made the study objectives and data protection guidelines transparent to all participants, who provided written consent prior to data collection. We guaranteed the absence of nonparticipants during the focus groups. We audio recorded all focus groups and uploaded them to a secure server based at Heidelberg University Hospital, which was only accessible to the research team. Additionally, we made field notes.

Analysis

A professional transcription service completed verbatim audio transcriptions of all focus groups. We did not return transcripts to participants for comment. We pseudonymized all transcripts according to data protection guidelines. Two members of the research team (M Hoffmann and MWH) independently performed content analysis [21] using MAXQDA 2018, following a two-step, deductive-inductive, and data-driven development of a system of categories. First, we used the Tailored Implementation in Chronic Diseases (TICD) framework [22] as a coding framework to deductively identify relevant determinants, focusing on barriers and facilitators of the acceptance of mental health specialist video consultations. The TICD framework is regarded as particularly appropriate for processing data focusing on implementation factors in order to tailor an intervention to specific needs. Second, we inductively generated the subcodes based on the analysis of every transcript. Both researchers compared their analysis, and disagreements throughout the process were discussed and resolved among the research team. We analyzed the remaining transcripts based on the developed coding system. In order to ensure that all key aspects were represented in the coding system, codes were consequently modified when new aspects emerged. Table MA3-1 in Multimedia Appendix 3 provides a summary of the coding system. We did not ask participants to provide feedback on the findings.

Results

Overview

Table 1 shows the characteristics of the 11 participants. Specifically, we interviewed 4 licensed psychological psychotherapists (36%), 3 physicians with board certification in psychosomatic medicine and psychotherapy (27%), 3 board-certified psychiatrists (27%), and 1 medical doctor with an additional qualification in psychotherapy (psychotherapeutisch tätiger Arzt, in German) (9%).

Depiction of Current Practice

Three main topics related to the current practice of patients with mental disorders in outpatient care emerged. First, long waiting times and an insufficient number of available mental health specialists (10/11, 91%) were the most frequently discussed issue, particularly due to the urban-rural divide. Especially for more severely affected patients, participants perceived long travel distances as well as long waiting times as barriers for the uptake of psychotherapy.

> *Sometimes the patients must wait for a very long time and sometimes they are also very burdened. They are also not able to call five other psychotherapists and drive to all of them.* [Participant #3, focus group #2]

Second, mental health specialists identified another health care gap regarding some patient groups and their difficulties in finding a therapist (6/11, 55%), such as the elderly, patients with impaired mobility, as well as severely affected patients (eg, those with borderline personality disorders). Third, beyond patient-related factors, mental health specialists perceived a lack of collaboration with family physicians (6/11, 55%), for instance, the absence of comprehensive referral letters (4/11, 36%) and the lack of an opportunity to consult with specialists in cases of diagnostic uncertainty (5/11, 45%). This was often associated with limited time resources of the mental health specialist and the family physician.

Table 1. Characteristics of the sample participants.

Characteristic	Value (N=11)
Sex (male), n (%)	2 (18)
Age in years (n=10), mean (SD)	52.3 (12.8)
Years in office-based practice (n=8), mean (SD)	13.1 (11.8)
Areas of recruitment, n (%)	
Cities (ie, densely populated areas)	6 (55)
Towns and suburbs (ie, intermediate-density areas) and rural areas (ie, thinly populated areas)	5 (45)
Therapeutic approach, n (%)	
Psychodynamic psychotherapy	5 (45)
Psychoanalysis	3 (27)
Behavior therapy	3 (27)

Factors Determining the Acceptance of Mental Health Specialist Video Consultations

We present barriers and facilitators regarding the acceptance of mental health specialist video consultations in a descriptive manner alongside the seven TICD framework domains in descending order by code frequency. Two domains—namely, individual health professional factors and patient factors—were identified as highly relevant and most frequently discussed. Some domains—namely, social, political, and legal factors as well as guideline factors—were hardly discussed and are not reported.

Individual Health Professional Factors

Regarding the acceptance of the intervention model, mental health specialists most frequently discussed individual health professional factors (11/11, 100%). For instance, they perceived some organizational factors as potential challenges.

> *Of course, there are organizational issues which need to be addressed. Primarily, how can it be managed without any disadvantages for me...I still do not have more capacities this way.* [Participant #1, focus group #2]

According to most of the participants (9/11, 82%), time and duration of the individual video consultation may be one major challenge of implementing the intervention. For example, some participants perceived a video consultation to be more exhaustive compared to a face-to-face appointment and, therefore, considered 25 minutes to be an appropriate duration. Others emphasized that mental health specialist video consultations should last as long as face-to-face consultations, which usually run for 50 minutes. They argued that this would contribute to establishing a therapeutic setting of the video consultation and, therefore, 50 minutes would provide a sufficient duration to get a comprehensive understanding of the patients' situations. When asked about potential benefits or expected outcomes, participants commonly highlighted the collaboration with the family physician (6/11, 55%) and help for patients (7/11, 64%). Overall, mental health specialists expected fewer advantages for themselves than for patients, especially compared to their everyday practice in the face-to-face setting. As a major issue, which has been subject to every focus group, mental health specialists questioned whether video consultations can be an appropriate and suitable means to deliver specialized mental health care to patients in need (11/11, 100%).

> *So, the exclusive situation, precisely that you do not evade...this gets somewhat socially lost as a fundamental, therapeutic experience.* [Participant #3, focus group #3]

Mental health specialists most commonly highlighted the advantages of the traditional face-to-face setting, while emphasizing some disadvantages of mental health specialist video consultations. Specifically, they underscored a lack of personal interactions and nonverbal cues, which may impede the therapeutic relationship and engagement (8/11, 73%).

> *This may also entail several disadvantages for me. Because it is a different setting, sitting together in a room...I believe the ritual of doing therapy together in a room is too important.* [Participant #1, focus group #2]

Some participants considered the model to work better when the patients already had previous appointments with the therapist and when a sustainable patient-therapist relationship already existed (3/11, 27%). Others (4/11, 36%) expected the model not to work as therapy in the traditional sense. Moreover, they emphasized that mental health specialist video consultations might dehumanize the therapist-patient relationship and, consequently, should not replace face-to-face therapies. Against this background, mental health specialists considered it to be essential to clearly define the purpose of the intervention and how it can make a significant contribution to improve the patients' situations.

> *I believe it is important to define the framework. What is supposed to happen? Psychotherapeutic video consultation implies the message that something therapeutic happens. But we all shared to some extent the opinion that it is not likely that something psychotherapeutic can really happen.* [Participant #3, focus group #2]

From the perspective of the mental health specialists, the model may be particularly suited for an initial mental health consultation. Within this setting, mental health specialists could serve as experts to provide more rapid access to mental health care. Some participants stated that their professional and therapeutic habits, as well as their daily routines to work in a face-to-face setting, might lead to a more conservative attitude toward digital health interventions, in general, and mental health specialist video consultations, in particular (5/11, 45%). One mental health specialist stated the following:

> *And getting used to this, yes. That is certainly a difference. Now with the monitor and talking to the patient in this way. The familiarization. But I think it is also exciting.* [Participant #1, focus group #1]

Using the intervention in practice could reduce reservations and enable mental health specialists to become familiar with the video consultation.

Patient Factors

First, within this domain, mental health specialists most frequently discussed appropriate target groups for the intervention (10/11, 91%). They considered mental health specialist video consultations to be most suitable for patients with impaired mobility (6/11, 55%) who may otherwise remain untreated. The following statement by another participant emphasized that the model could serve as an alternative model to deliver mental health care, especially in rural areas.

> *...this is not supposed to replace conventional psychotherapy; rather, just for particular cases, right? For people who just live too far away, where the infrastructure is very, very poor.* [Participant #1, focus group #1]

In contrast, some participants mentioned patients with specific mental health conditions, such as anxiety disorders, or patients with posttraumatic stress disorders (3/11, 27%) as potentially less suitable for mental health specialist video consultations. Hence, they stated that it should be clearly specified which patients are the main target group to tailor the intervention to their needs. Furthermore, some mental health specialists indicated that in the event of an emergency, it should be clearly defined who is on site in general practice and responsible for the patient (3/11, 27%). Second, regarding patient barriers, participants anticipated that some patients might have concerns about the video technology, as such (5/11, 45%). This might lead to less interest or refusal to conduct mental health specialist video consultations, especially regarding the elderly. Therefore, it might take some extra effort to encourage the acceptance of mental health specialist video consultations. Considering potential benefits for patients, they valued quicker help, low-threshold access, and the initiation of psychotherapy as main facilitators.

> *I could very well imagine doing something like this. To offer a low-threshold service, to initiate psychotherapy, very carefully and sensitively.* [Participant #4, focus group #2]

As a potential result, long waiting times, a tedious search for an appointment, as well as long travel distances could be avoided.

Professional Interactions

Against the background of a lack of collaborative exchange with family physicians as outlined above, mental health specialists expected professional interactions with the family physician to be a potential benefit from, and facilitator for, mental health specialist video consultations (7/11, 64%). Some participants considered the model to foster collaboration with the family physician and highly valued the possibility to give feedback to the family physician and better involve him or her in the process of care (5/11, 45%). Specifically, they considered brief case discussions or the virtual handing over of the patient by the physician as advantageous.

> *For the initial contact, it would certainly be nice if the general practitioner would facilitate it by introducing it [the model] and state what he has thought of. To already sort of point me in the right direction...And then I could, I could go deeper with the patient. And maybe at the end, I could share the result, with him [family physician]...What, from my view, I believe could, should be done.* [Participant #1, focus group #2]

Some participants also discussed the possibility of a consultation not only between the patient and the mental health specialist but also with the family physician, who could attend the video consultation (3/11, 27%). In this setting, mental health specialists would want to discuss the patient's condition with the patient and the family physician. As a result, this could foster collaboration between the family physician and the mental health specialist, as well as the patient's involvement.

> *I think of a trialogue, for example, between the general practitioner, the patient, and psychotherapist and one asks the psychotherapist for his suggestion.* [Participant #4, focus group #3]
>
> *Yes, I also find the trialogue very exciting, because this way the patient and the general practitioner are on a triangulating level with me as a consultant. This also changes the relationship between them and adds something new.* [Participant #1, focus group #3]

Furthermore, within this domain, some mental health specialists discussed the distribution of roles and responsibilities (4/11, 36%). First, regarding family physicians' roles, referring patients would be an essential task. Also, as described above, participants considered as potentially helpful the introduction of the patient to the mental health specialist at the beginning of the video consultation. Second, mental health specialists regarded themselves as expert consultants, especially against the background of video consultations as a different model of delivery compared to traditional face-to-face psychotherapy.

> *And this would possibly also be the framework; for the patient, the psychotherapist would be the expert, so to say...Sort of, consulting someone and to get clarification in a particular, temporarily defined setting, and where the patient could possibly save trips.* [Participant #3, focus group #3]

Incentives and Resources

First, within this domain, the majority of participants extensively discussed costs and payment conditions related to the intervention (7/11, 64%). As outlined in the first domain, participants do see more disadvantages than benefits, especially compared to their daily face-to-face working practice. Therefore, mental health specialists regarded a prerequisite of a reasonable financial remuneration to be an incentive to conducting mental health specialist video consultations themselves.

> *I believe it is important to settle financial issues. Particularly, because therapists are not at all dependent on this [the model].* [Participant #1, focus group #1]

Some participants stated that the model should include remuneration of video consultations to be equal to face-to-face sessions (3/11, 27%) or higher (2/11, 18%), in order to enhance the mental health specialists' motivation to conduct video consultations. Also, considering financial compensation, 2 participants (18%) stated that the video consultation sessions should be remunerated even when patients were not able to keep the appointment. Otherwise, this might negatively affect mental health specialists' acceptance of the treatment model, as mental health specialists in Germany receive a payment in cases where the patient misses an appointment because most mental health specialists run an appointment-based practice. In addition to that, they underscored stable network connectivity, high visual definition, minimized speech delay, minimized interruptions, and the preinstallation of equipment as essential. Technical equipment should be provided and not be paid for by the mental health specialists themselves. Considering the coordination of appointments, participants stated that fixed time slots are a prerequisite as well as the duty of primary care practices to efficiently implement the intervention into their daily working practice. Overall, participants required organizational demands to be settled in order to commit themselves to mental health specialist video consultations, as illustrated by the following quote:

> *If we do not consider all organizational issues, as how is it practically possible...if I look at it from a completely IDEALISTIC perspective, then this would not be a bad idea. Let´s say as an initial contact.* [Participant #1, focus group #2]

Capacity for Organizational Change

Mental health specialists discussed aspects regarding the capacity for organizational change only when explicitly asked about their perceptions of the readiness for change of mental health specialists in general. Some participants expected heterogeneous attitudes among mental health specialists in general. Others expected younger mental health specialists to be more curious and open minded (2/11, 18%).

> *The people above 50 might be a little bit more reluctant; about the younger ones, I am not entirely sure. But I could imagine that they are more adventurous.* [Participant #2, focus group #1]

One participant suspected that either intrinsic motivation—namely, providing help for patients—or financial incentives would be a potential means to drive the capacity for organizational change.

> *I could imagine that those colleagues would agree to this [the model], who do this as additional benefit for the patient. Or one would have to compensate it in a way that it becomes financially profitable.* [Participant #5, focus group #2]

Differences in Views in Relation to Urbanization and Therapeutic Orientation

To account for potential distinctions, we compared data from cities (6/11, 55%) to that of suburbs and rural areas (5/11, 45%), as well as different therapeutic orientations—namely, psychodynamic psychotherapy and psychoanalysis psychotherapy (8/11, 73%) as well as cognitive behavioral therapy (3/10, 27%). First, a comparison of the degree of urbanization revealed that participants from cities (4/6, 67%) more frequently mentioned the appointment time and scheduling of the mental health specialist video consultations and technical requirements as prerequisites compared to participants from less-urbanized areas. Also, mental health specialists from cities (4/6, 67%) more often highlighted the possibility to collaborate with the family physician compared to specialists from suburbs or rural areas (2/5, 40%). Second, regarding different therapeutic approaches, we found that cognitive behavioral therapy–oriented participants considered clearly defined responsibilities in the event of an emergency (3/3, 100%), whereas participants with other therapeutic orientations did not discuss this issue. Also, cognitive behavioral therapy–oriented participants more frequently considered technical aspects and an impersonal setting within the mental health specialist video consultations model as potential barriers for some patients (3/3, 100%), compared to specialists practicing psychodynamic psychotherapy and psychoanalysis psychotherapy (4/8, 50%). Taking into consideration that all participants underscored the importance of physical presence for psychotherapy, mental health specialists with a background in psychodynamic psychotherapy and psychoanalysis psychotherapy particularly mentioned patients with impaired mobility as an appropriate target group for the model (5/8, 63%). Cognitive behavioral therapy–oriented participants (1/3, 33%) indicated these patients as a target group less frequently.

Discussion

Principal Findings

The aim of this study was to explore how mental health specialists perceive the acceptance of mental health specialist video consultation embedded into primary care practice against the background of their perception of current mental health care. Most mental health specialists described the latter as insufficient due to long waiting times and a lack of therapeutic capacities, especially in rural areas. We explored crucial barriers and facilitators of mental health specialist video consultations as a potential mode to deliver mental health care. First and foremost, mental health specialists questioned whether a stable therapeutic alliance can be established, especially with respect to the therapeutic alliance and engagement. Therefore, they considered mental health specialist video consultations to be particularly suited for an initial consultation and diagnosis, rather than a genuine alternative to face-to-face psychotherapy. Second, regarding mental health specialists' perceptions of the patient perspective, they highlighted fast access to mental health care, especially for patients with impaired mobility, as a major benefit of the model. Participants considered mental health treatment via video consultations not to be a barrier but to be rather predominantly advantageous for most patients. Third, apart from an intrinsic motivation—namely, to provide faster access to mental health care—mental health specialists considered financial incentives as instrumental. Fourth, participants underscored important organizational factors, such as stable network connectivity, high visual definition, and minimized interruptions. Therefore, training sessions in order to familiarize users with the video consultation in practice are key drivers for the intervention.

Strengths and Limitations

Our study has some limitations. First, due to the small sample size and only one mental health specialist from rural areas participating in our study, the results and distinctions between participants from different areas are reliable to a limited degree. Second, none of the participants had conducted mental health specialist video consultations before participating in the study, since we conducted the focus groups prior to implementation of the proposed intervention model. It is plausible that some mental health specialists would revise their initial perceptions regarding the intervention model after the actual implementation. However, by assessing the perceptions before the implementation, we explored necessary prerequisites for further tailoring mental health specialist video consultations to the mental health specialists' needs, which is key for the adoption of a new mode of delivering mental health care. Third, we did not apply a technology acceptance framework (eg, technology acceptance model [TAM] or unified theory of acceptance and use of technology [UTAUT]). Nevertheless, while such models support the explanation of the potential user's behavioral intention to engage with a technological innovation [23], our focus was on meso-level implementation in a broader sense. Thus, we employed the TICD framework for guiding both the analysis of our data and interpretation of our findings. Finally, not all barriers related to the proposed model may have been referred to by the study participants. For example, it is plausible that the distance between the patient's home and the general practice itself could be a problem for patients struggling with limited mobility. However, we would argue that patients usually need to see their general practitioner eventually for regular checkups and/or the monitoring and treatment of medical diseases. From a patient-centered perspective, the proposed video consultations could be included in these appointments. At the very least, patients would save time and avoid travel related to additional in-person visits with mental health specialists.

Comparison With Prior Work

Previous work on video consultations, and telemedicine in general, has mainly focused on efficiency trials, satisfaction, and attitudes after the implementation of video consultations. A large body of research points to a comparable effectiveness of video consultations and face-to-face consultations [4,24-27]. Referring to the patients' perspective, many studies found high satisfaction rates with video consultations for addressing various health conditions [4,9,27,28]. In contrast, another study found higher satisfaction rates for patients than for clinical staff [27]. However, within this work (ie, focusing on chronic health conditions), reasons for the lower level of satisfaction among clinical staff remained unclear. Our results add that mental health specialists were most often concerned about a potential dehumanization of the therapeutic alliance caused by a lack of personal interactions and nonverbal cues. Another online survey study among stakeholders in eight European countries found that care providers had particular concerns regarding the therapeutic alliance [29], which is in line with our findings. However, we provide additional insights into the perspective of mental health specialists, as we also investigated necessary prerequisites and key drivers, such as the collaboration with the general practitioner, specific target groups, and the virtual handing over of the patient by the physician. A systematic review on the therapeutic alliance in videoconferencing found that therapists rated the therapeutic alliance as high, although not as high as their clients did, but often increasing over the course of therapy [30].

Our study, as well as other studies, point to the importance of training for mental health specialists regarding the use of video consultations [30-34]. Training sessions should impart knowledge regarding the preparation and technical conduct of video consultations as well as encourage the therapist's ability for reflection. This might potentially reduce some reservations.

Additionally, we found that training sessions should especially focus on establishing a therapeutic alliance via video, such as ways to greet the patient virtually and open the session, and on fostering therapeutic and patient engagement as crucial preconditions to adopt mental health specialist video consultations. We found that mental health specialists evaluated the acceptance of mental health specialist video consultations almost entirely in comparison to face-to-face consultations and particularly emphasized the benefits and qualities of the latter. Specifically, we found that participants in our study mainly did not consider mental health specialist video consultations as a favorable alternative mode to delivering mental health care, which some mental health specialists associated with their professional habits and daily routines. Exceptions were mental health specialist video consultations for otherwise untreated patients, especially in remote and rural areas, and mental health specialist video consultations as initial consultations. Besides training, previous studies also revealed that an existing relationship between the patient and the health care provider is crucial [35,36], which some mental health specialists in our study also indicated. When it comes to the acceptance and adoption of new innovations, such as mental health specialist video consultations, the perceived compatibility of innovation with the values and norms of the potential adopters are key elements [37]. Therefore, mental health care delivered via video consultations should focus on these target groups in order to increase the compatibility of mental health specialist video consultations with the norms of mental health specialists and their daily practices. Previous studies on this topic have been limited to perceptions after the treatment model has already been implemented. Regarding our previous work on family physicians' perceptions of mental health specialist video consultations in primary care [14], this study adds necessary insights from the perspective of mental health specialists on the proposed treatment model. Comparable to the main barrier perceived by mental health specialists, which is the therapeutic alliance, family physicians in our previous study underscored a lack of personal interactions and nonverbal cues that may impede therapeutic engagement. Also, they valued benefits not only for patients but also for the family physicians themselves—namely, that mental health specialist video consultations may free up resources for the primary care practice. However, in this study, we found that mental health specialists perceived more disadvantages for themselves and their therapeutic work, as described above.

Conclusions

To increase access to specialized care, emergent technologies like interactive video consultations, which are known to be highly accepted by patients, will probably complement traditional in-person consultations in future health care systems. However, we found that mental health specialists are still somewhat skeptical concerning the impact on the patient-therapist relationship. To foster the implementation and adoption of mental health specialist video consultations in primary care, it is essential to train providers in managing technology-based treatment settings with extra consideration for preserving therapeutic alliance and to familiarize mental health specialists with video consultations. Based on the findings from this study and our previous work, we started a feasibility study (trial registration number: DRKS00015812) to further tailor the intervention model to the participants' needs and to evaluate the acceptance and practicability of mental health specialist video consultation in primary care [15].

Acknowledgments

This research was supported by grants from the Federal Ministry of Education and Research (BMBF) (grant No. 01GY1612). The funder was not actively involved in the conduct of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

COREQ (COnsolidated criteria for REporting Qualitative research) checklist.

[PDF File (Adobe PDF File), 425 KB - jmir_v22i6e17569_app1.pdf]

Multimedia Appendix 2

Semistructured guide for focus groups and telephone interview.

[DOCX File , 16 KB - jmir_v22i6e17569_app2.docx]

Multimedia Appendix 3

Description of the coding system.

[DOCX File , 19 KB - jmir_v22i6e17569_app3.docx]

References

1. Fortney JC, Pyne JM, Turner EE, Farris KM, Normoyle TM, Avery MD, et al. Telepsychiatry integration of mental health services into rural primary care settings. Int Rev Psychiatry 2015;27(6):525-539. [doi: 10.3109/09540261.2015.1085838] [Medline: 26634618]

2. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders. J Clin Psychiatry 2013 Jun 15;74(06):595-602. [doi: 10.4088/jcp.12r07757]
3. Vanderlip ER, Rundell J, Avery M, Alter C, Engel C, Fortney J, et al. Dissemination of Integrated Care Within Adult Primary Care Settings: The Collaborative Care Model. Washington, DC: American Psychiatric Association and Academy of Psychosomatic Medicine; 2016. URL: https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/Professional-Topics/Integrated-Care/APA-APM-Dissemination-Integrated-Care-Report.pdf [accessed 2020-06-01]
4. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: A systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]
5. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, et al. Videoconferencing psychotherapy and depression: A systematic review. Telemed J E Health 2019 Jun;25(6):435-446. [doi: 10.1089/tmj.2018.0058] [Medline: 30048211]
6. Chakrabarti S. Usefulness of telepsychiatry: A critical evaluation of videoconferencing-based approaches. World J Psychiatry 2015 Sep 22;5(3):286-304 [FREE Full text] [doi: 10.5498/wjp.v5.i3.286] [Medline: 26425443]
7. Hilty DM, Rabinowitz T, McCarron RM, Katzelnick DJ, Chang T, Bauer AM, et al. An update on telepsychiatry and how it can leverage collaborative, stepped, and integrated services to primary care. Psychosomatics 2018;59(3):227-250. [doi: 10.1016/j.psym.2017.12.005] [Medline: 29544663]
8. Richardson LK, Frueh BC, Grubaugh AL, Egede L, Elhai JD. Current directions in videoconferencing tele-mental health research. Clin Psychol (New York) 2009 Sep 01;16(3):323-338 [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01170.x] [Medline: 20161010]
9. Stubbings DR, Rees CS, Roberts LD, Kane RT. Comparing in-person to videoconference-based cognitive behavioral therapy for mood and anxiety disorders: Randomized controlled trial. J Med Internet Res 2013 Nov 19;15(11):e258 [FREE Full text] [doi: 10.2196/jmir.2564] [Medline: 24252663]
10. Yellowlees P, Shore JH. Telepsychiatry and Health Technologies: A Guide for Mental Health Professionals. Arlington, VA: American Psychiatric Association Publishing; 2018.
11. Fortney JC, Pyne JM, Edlund MJ, Williams DK, Robinson DE, Mittal D, et al. A randomized trial of telemedicine-based collaborative care for depression. J Gen Intern Med 2007 Aug;22(8):1086-1093 [FREE Full text] [doi: 10.1007/s11606-007-0201-9] [Medline: 17492326]
12. Simpson S. Psychotherapy via videoconferencing: A review. Br J Guid Counc 2009 Aug;37(3):271-286. [doi: 10.1080/03069880902957007]
13. Grol R. Personal paper. Beliefs and evidence in changing clinical practice. BMJ 1997 Aug 16;315(7105):418-421 [FREE Full text] [doi: 10.1136/bmj.315.7105.418] [Medline: 9277610]
14. Hoffmann M, Hartmann M, Wensing M, Friederich H, Haun MW. Potential for integrating mental health specialist video consultations in office-based routine primary care: Cross-sectional qualitative study among family physicians. J Med Internet Res 2019 Aug 19;21(8):e13382 [FREE Full text] [doi: 10.2196/13382] [Medline: 31429419]
15. Tönnies J, Hartmann M, Wensing M, Szecsenyi J, Icks A, Friederich H, et al. Mental health specialist video consultations for patients with depression or anxiety disorders in primary care: Protocol for a randomised controlled feasibility trial. BMJ Open 2019 Sep 04;9(9):e030003 [FREE Full text] [doi: 10.1136/bmjopen-2019-030003] [Medline: 31488484]
16. Morgan DL. Focus groups. Annu Rev Sociol 1996 Aug;22(1):129-152. [doi: 10.1146/annurev.soc.22.1.129]
17. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
18. Palys T. Purposive sampling. In: Given LM, editor. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, CA: SAGE Publications; 2008:697-698.
19. Dijkstra L, Poelman H. A Harmonised Definition of Cities and Rural Areas: The New Degree of Urbanisation. Brussels, Belgium: European Commission; 2014. URL: https://ec.europa.eu/regional_policy/sources/docgener/work/2014_01_new_urban.pdf [accessed 2020-06-01]
20. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research. 5th edition. Thousand Oaks, CA: SAGE Publications; 2015:1-252.
21. Kuckartz U. Qualitative Text Analysis: A Guide to Methods, Practice & Using Software. Thousand Oaks, CA: SAGE Publications; 2014:1-173.
22. Flottorp SA, Oxman AD, Krause J, Musila NR, Wensing M, Godycki-Cwirko M, et al. A checklist for identifying determinants of practice: A systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implement Sci 2013 Mar 23;8:35 [FREE Full text] [doi: 10.1186/1748-5908-8-35] [Medline: 23522377]
23. King WR, He J. A meta-analysis of the technology acceptance model. Inf Manag 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]
24. Day SX, Schneider PL. Psychotherapy using distance technology: A comparison of face-to-face, video, and audio treatment. J Couns Psychol 2002;49(4):499-503. [doi: 10.1037/0022-0167.49.4.499]

25. Lamb T, Pachana NA, Dissanayaka N. Update of recent literature on remotely delivered psychotherapy interventions for anxiety and depression. Telemed J E Health 2019 Aug;25(8):671-677. [doi: 10.1089/tmj.2018.0079] [Medline: 30300082]
26. Norwood C, Moghaddam NG, Malins S, Sabin-Farrell R. Working alliance and outcome effectiveness in videoconferencing psychotherapy: A systematic review and noninferiority meta-analysis. Clin Psychol Psychother 2018 Nov;25(6):797-808. [doi: 10.1002/cpp.2315] [Medline: 30014606]
27. Steel K, Cox D, Garry H. Therapeutic videoconferencing interventions for the treatment of long-term conditions. J Telemed Telecare 2011;17(3):109-117. [doi: 10.1258/jtt.2010.100318] [Medline: 21339304]
28. Richardson L, Reid C, Dziurawiec S. “Going the extra mile”: Satisfaction and alliance findings from an evaluation of videoconferencing telepsychology in rural Western Australia. Aust Psychol 2015 Jul 20;50(4):252-258. [doi: 10.1111/ap.12126]
29. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED Consortium. Attitudes towards digital treatment for depression: A European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
30. Simpson SG, Reid CL. Therapeutic alliance in videoconferencing psychotherapy: A review. Aust J Rural Health 2014 Dec;22(6):280-299. [doi: 10.1111/ajr.12149] [Medline: 25495622]
31. Clay-Williams R, Baysari M, Taylor N, Zalitis D, Georgiou A, Robinson M, et al. Service provider perceptions of transitioning from audio to video capability in a telehealth system: A qualitative evaluation. BMC Health Serv Res 2017 Aug 14;17(1):558 [FREE Full text] [doi: 10.1186/s12913-017-2514-7] [Medline: 28806903]
32. Hilty DM, Randhawa K, Maheu MM, McKean AJS, Pantera R. Therapeutic relationship of telepsychiatry and telebehavioral health: Ideas from research on telepresence, virtual reality and augmented reality. Psychol Cogn Sci 2019 Aug 26;5(1):14-29 [FREE Full text] [doi: 10.17140/pcsoj-5-145]
33. Mozer R, Bradford NK, Caffery LJ, Smith AC. Identifying perceived barriers to videoconferencing by rehabilitation medicine providers. J Telemed Telecare 2015 Dec;21(8):479-484. [doi: 10.1177/1357633X15607136] [Medline: 26556061]
34. Titov N, Hadjistavropoulos HD, Nielssen O, Mohr DC, Andersson G, Dear BF. From research to practice: Ten lessons in delivering digital mental health services. J Clin Med 2019 Aug 17;8(8):1239 [FREE Full text] [doi: 10.3390/jcm8081239] [Medline: 31426460]
35. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: A qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
36. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: Mixed method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
37. Rogers EM. Diffusion of Innovations. 5th edition. New York, NY: Free Press; 2003:1-551.

Abbreviations

BMBF: Federal Ministry of Education and Research
COREQ: COnsolidated criteria for REporting Qualitative research
DEGURBA: Degree of Urbanisation
PROVIDE: ImPROving cross-sectoral collaboration between primary and psychosocial care: An implementation study on VIDEo consultations
RCT: randomized controlled trial
TAM: technology acceptance model
TICD: Tailored Implementation in Chronic Diseases
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by N Vandenberg, C Jacob; comments to author 10.03.20; revised version received 18.03.20; accepted 03.04.20; published 18.06.20.

Please cite as:
Hoffmann M, Wensing M, Peters-Klimm F, Szecsenyi J, Hartmann M, Friederich HC, Haun MW
Perspectives of Psychotherapists and Psychiatrists on Mental Health Care Integration Within Primary Care Via Video Consultations: Qualitative Preimplementation Study
J Med Internet Res 2020;22(6):e17569
URL: http://www.jmir.org/2020/6/e17569/
doi:10.2196/17569
PMID:32554369

Original Paper

Improvements in Patient Monitoring in the Intensive Care Unit: Survey Study

Akira-Sebastian Poncette[1,2], MD; Lina Mosch[1]; Claudia Spies[1], MD; Malte Schmieding[1,2], MD; Fridtjof Schiefenhövel[1,2], MD; Henning Krampe[1], PhD; Felix Balzer[1,2], MD, MSc, PhD

[1]Department of Anesthesiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[2]Einstein Center Digital Future, Berlin, Germany

Corresponding Author:
Felix Balzer, MD, MSc, PhD
Department of Anesthesiology and Intensive Care Medicine
Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health
Charitéplatz 1
Berlin
Germany
Phone: 49 30 450 651166
Email: felix.balzer@charite.de

Abstract

Background: Due to demographic change and, more recently, coronavirus disease (COVID-19), the importance of modern intensive care units (ICU) is becoming apparent. One of the key components of an ICU is the continuous monitoring of patients' vital parameters. However, existing advances in informatics, signal processing, or engineering that could alleviate the burden on ICUs have not yet been applied. This could be due to the lack of user involvement in research and development.

Objective: This study focused on the satisfaction of ICU staff with current patient monitoring and their suggestions for future improvements. We aimed to identify aspects of monitoring that interrupt patient care, display devices for remote monitoring, use cases for artificial intelligence (AI), and whether ICU staff members are willing to improve their digital literacy or contribute to the improvement of patient monitoring. We further aimed to identify differences in the responses of different professional groups.

Methods: This survey study was performed with ICU staff from 4 ICUs of a German university hospital between November 2019 and January 2020. We developed a web-based 36-item survey questionnaire, by analyzing a preceding qualitative interview study with ICU staff, about the clinical requirements of future patient monitoring. Statistical analyses of questionnaire results included median values with their bootstrapped 95% confidence intervals, and chi-square tests to compare the distributions of item responses of the professional groups.

Results: In total, 86 of the 270 ICU physicians and nurses completed the survey questionnaire. The majority stated they felt confident using the patient monitoring equipment, but that high rates of false-positive alarms and the many sensor cables interrupted patient care. Regarding future improvements, respondents asked for wireless sensors, a reduction in the number of false-positive alarms, and hospital standard operating procedures for alarm management. Responses to the display devices proposed for remote patient monitoring were divided. Most respondents indicated it would be useful for earlier alerting or when they were responsible for multiple wards. AI for ICUs would be useful for early detection of complications and an increased risk of mortality; in addition, the AI could propose guidelines for therapy and diagnostics. Transparency, interoperability, usability, and staff training were essential to promote the use of AI. The majority wanted to learn more about new technologies for the ICU and required more time for learning. Physicians had fewer reservations than nurses about AI-based intelligent alarm management and using mobile phones for remote monitoring.

Conclusions: This survey study of ICU staff revealed key improvements for patient monitoring in intensive care medicine. Hospital providers and medical device manufacturers should focus on reducing false alarms, implementing hospital alarm standard operating procedures, introducing wireless sensors, preparing for the use of AI, and enhancing the digital literacy of ICU staff. Our results may contribute to the user-centered transfer of digital technologies into practice to alleviate challenges in intensive care medicine.

Trial Registration: ClinicalTrials.gov NCT03514173; https://clinicaltrials.gov/ct2/show/NCT03514173

(J Med Internet Res 2020;22(6):e19091) doi:10.2196/19091

KEYWORDS

digital health; patient monitoring; monitoring; intensive care medicine; intensive care unit; technological innovation; user-centered; usability; online survey; transdisciplinary; REDCap; email

Introduction

Background

In the near future, continuous monitoring of patients' vital signs will play an increasingly important role in alleviating the burden on the health care system caused by demographic change and, more recently, coronavirus disease (COVID-19) [1]. Both lead to an increased number of critically ill patients requiring intensive medical care, including mechanical ventilation and patient monitoring. However, existing advances in informatics, signal processing, or engineering have not yet been applied to patient monitoring [2], making it primarily an alarm system notifying health care providers whenever a patient's parameter deviates from preset values that are considered safe. To accelerate technology transfer into clinical routine, it may be beneficial to include users' pain points and suggestions for research and development.

Patient monitoring can be applied across almost all health sectors, which underlines its importance and the potential offered by digitalization. First, patients can monitor themselves preventively (eg, for atrial fibrillation), even with a consumer product such as the Apple Watch [3]. Second, remote monitoring of patients over long distances is a crucial component of telemedicine, which is becoming increasingly widespread in most areas of medicine [4]. Third, patient monitoring might soon be mandatory in general wards due to a shift in inpatient clientele toward the more critically ill [5,6]. Finally, patient monitoring produces high-frequency data that are a valid and essential source for clinical decision support systems (CDSS) based on artificial intelligence (AI), opening up many possibilities for precision medicine [7].

In the intensive care unit (ICU), as one of the most technologically enhanced medical areas, staff have used monitoring technologies over decades. In a previous qualitative study from our research group, ICU staff demanded wireless, noninvasive, and interoperable monitoring sensors and improved alarm management for a future patient monitoring system [8]. Mobile phones were desired as displays for remote patient monitoring, and CDSS based on AI was considered useful. To validate these inclinations in a larger cohort, we designed this survey study of ICU staff.

Aim

This survey study focuses on ICU staff members’ satisfaction with the current patient monitoring system and their suggestions for future technological improvements. In particular, we aimed to identify the aspects of patient monitoring that disturb patient care, the display devices most appropriate for the ICU for remote patient monitoring on the hospital premises, the use cases for AI in the ICU, and whether ICU staff is willing to improve their digital literacy or contribute to product improvement. With regard to the multiprofessional structure of ICU teams, we further desired to uncover differences in perspectives between different health professions in the ICU.

Methods

Ethics Approval and Consent to Participate

The ethical approval for this study was granted by the Ethics Commission of the Charité – Universitätsmedizin Berlin (EA1/031/18). Participation in the survey was voluntary. Prior to the study, all participants provided their written consent.

Setting

This survey study was performed with ICU staff from 4 ICUs of a German university hospital, between November 2019 and January 2020 as a substudy for the implementation of the virtual patient monitoring platform Vital Sync 2.4 (Medtronic plc). This new system was implemented between May 2018 and June 2019 in one of the 4 ICUs as a secondary patient monitoring system to remotely monitor patients via tablet computers. As the primary patient monitoring system, the Philips IntelliVue patient monitoring system (Koninklijke Philips NV; MX800 software version M.00.03; MMS X2 software version H.15.41-M.00.04) was used in all 4 ICUs at the time of the study. COPRA 6 (COPRA System GmbH) was used as the patient data management system (PDMS).

Study Design

We chose a cross-sectional survey design, and developed a web-based questionnaire [9,10]. Survey item generation was initiated through the analysis of a preceding qualitative interview study with ICU staff about clinical requirements of future patient monitoring, and was saturated in focus group sessions within the research team [8]. Items were then grouped into topics, and 5 to 6 items per topic were anticipated. We chose a 5-point Likert-type scale as an ordinal response format, with the options “Strongly disagree” (score=1), “Disagree” (score=2), “Undecided” (score=3), “Agree” (score=4), and “Strongly agree” (score=5). In pretests with associated research colleagues, redundant items were eliminated without removing whole topics. Pilot testing was conducted face-to-face with experts from intensive care medicine, with a focus on the clarity, relevance, and arrangement of the items into topics as well as the usability of the web-based questionnaire. Experts also assessed content validity (ie, whether all aspects of the topic were accurately covered by the questionnaire) and clinical validity (ie, whether the questionnaire measured the intended research topic). The final questionnaire (Multimedia Appendix 1) contained 36 items grouped into 8 topics:

- ICU staff experience with the current patient monitoring system
- Aspects of patient monitoring that disturb patient care
- Improvements for future patient monitoring
- Suggestions for remote patient monitoring display devices
- Use cases for remote patient monitoring
- Use cases for CDSS based on AI
- Aspects that promote the usage of CDSS based on AI
- Attitude of ICU staff toward novel digital technology

Additionally, respondents indicated their age group, profession, and technical affinity. For the latter, we used the Affinity for Technology Interaction Short (ATI-S) scale [11] and reduced the options from a 6-point scale to a 5-point Likert-type scale due to usability issues. Other items in the questionnaire focused on alarm management, which was the subject of another study and is not reported here.

Data Collection

Data collection took place over a period of 2 months (November 2019 to January 2020) on an invitation basis. The sampling frame was defined as the 270 nurses and physicians working in the 4 ICUs the day before data collection began; in total, there were 177 nurses and 93 physicians. An email containing a detailed description of the study and the web address of the survey was sent to them. Study data were collected and managed using REDCap electronic data capture tools hosted at Charité – Universitätsmedizin Berlin [12,13].

To increase the survey response rate, participants were offered the opportunity to take part in a raffle to win a €50 (US $56.04) voucher for a train ticket after survey participation. Additionally, 2 reminder emails were sent to all participants 2 and 5 weeks after the initial email was sent. Finally, small handouts with a brief description of the study, the URL for the questionnaire, and a QR (quick response) code were given to ICU staff on site.

Data Analysis

We cleaned and analyzed the data with R (R Foundation for Statistical Computing) in combination with the packages tidyverse, psych, and sjPlots [14-17]. Inferential calculations were performed with the infer package [18]. For each of the 36 five-point items, we calculated the medians and their 95% bootstrap CIs by deploying a bootstrap resampling procedure as previously described [19,20]. For the bootstrap sampling distribution, we created 15,000 bootstrap samples per item. An item median was considered statistically significant when the 95% bootstrap confidence intervals of the median did not include 3, which indicates the response “Undecided.” To compare the distributions of item responses of physicians and nurses, we used chi-square tests. Here, a two-tailed *P* value <.05 was considered statistically significant.

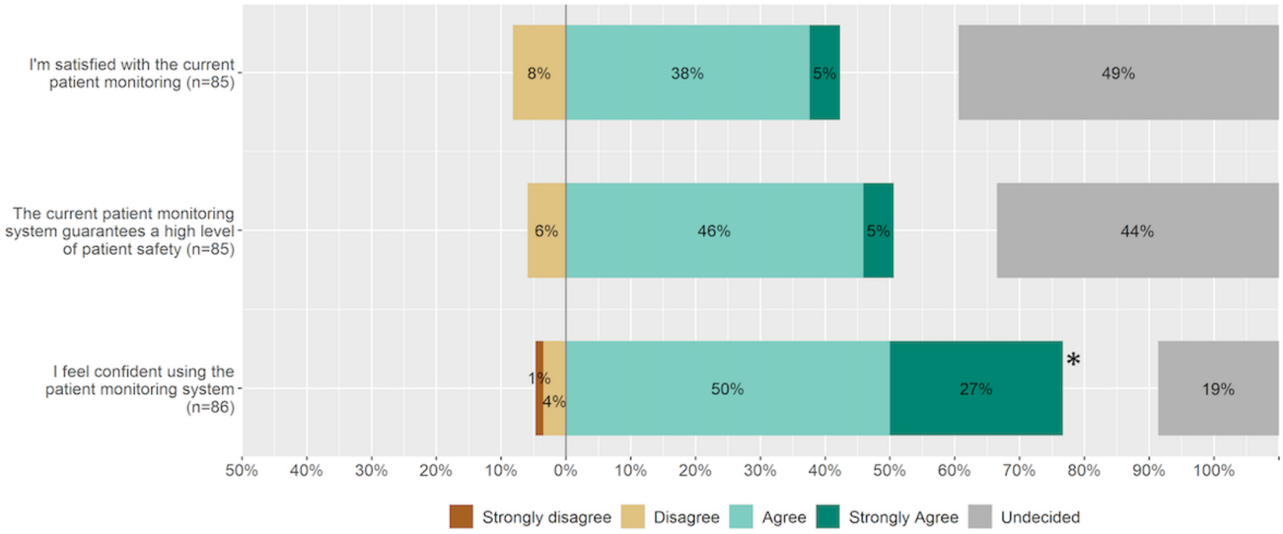
Results

Overview

This survey study is based on a questionnaire with 36 items regarding patient monitoring in the ICU, addressed to ICU staff. The actual response rate was 39.6% (107/270); however, only 86 responses from 62 nurses and 24 physicians were analyzable due to missing data. The ratio of male to female respondents was almost equal (42 men, 41 women, 3 not specified). The largest age categories were represented by participants aged 25 to 34 years (n=32, 37%) and those aged 35 to 44 years (n=28, 33%). Self-reported technical affinity (ATI-S) was rated with a mean of 3.4 (SD 0.88) and a median of 3.5 (range 2.9-4.1) on the 5-point Likert-type scale, with a Cronbach of 0.83 (95% CI 0.76-0.89).

The questionnaire results are presented as grouped Likert plots (Figures 1-8) [16], where one group represents one topic. An item median was considered statistically significant (items marked with an asterisk) when the 95% bootstrap CI of the median did not include 3, which indicates the response “Undecided” (Multimedia Appendix 2 shows item medians and bootstrap CIs). To improve readability, and in contrast to the questionnaire, the answer option “Undecided” is presented on the far right. Multimedia Appendix 3 contains the raw data, and Multimedia Appendix 4 shows the distribution of item responses of physicians and nurses.

Figure 1. ICU staff experience with current patient monitoring. An asterisk indicates statistical significance. ICU: intensive care unit.



Current Patient Monitoring

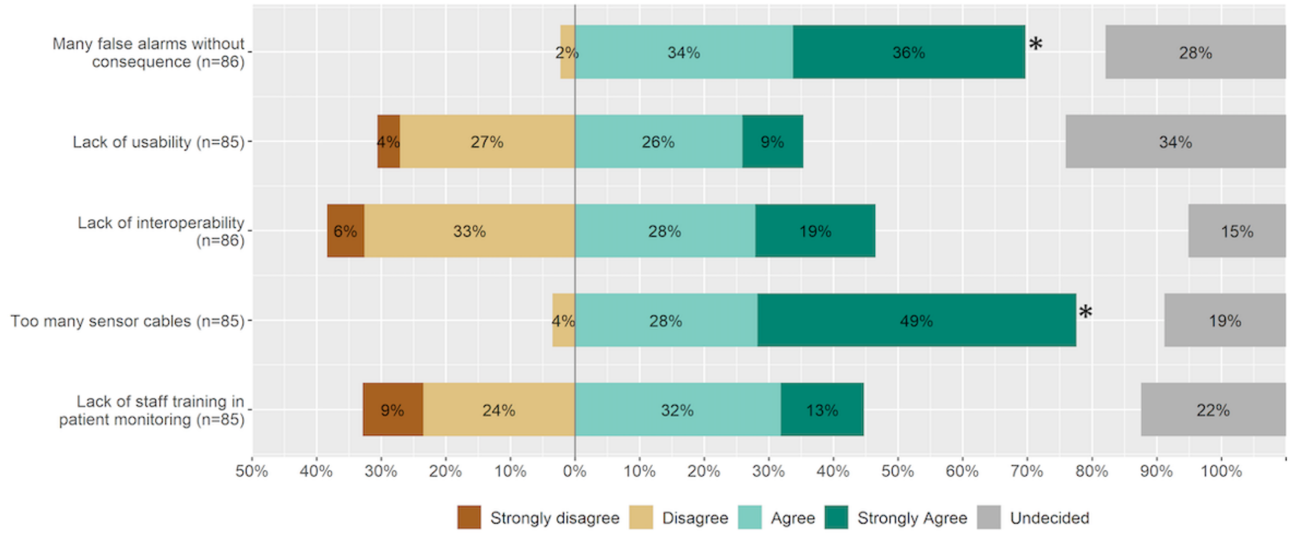
ICU Staff Experience

Most of the ICU staff who took part in the online survey were satisfied with the current patient monitoring system and felt that it ensured high patient safety, even though the median responses did not differ significantly from the option “Undecided” (Figure 1). The majority stated feeling confident in using the patient monitoring system (n=66, 77% chose “Strongly agree” or “Agree”).

Aspects Disturbing Patient Care

The majority of respondents indicated that the patient monitoring system’s high rate of false-positive alarms (n=60, 70% chose “Strongly agree” or “Agree”) and high number of sensor cables (n=66, 77% indicated “Strongly agree” or “Agree”) interrupted patient care. The opinions about detrimental effects elicited by a lack of interoperability, lack of staff training, and low usability of the patient monitoring system were split (Figure 2).

Figure 2. Aspects of patient monitoring disturbing patient care in the ICU. An asterisk indicates statistical significance. ICU: intensive care unit.

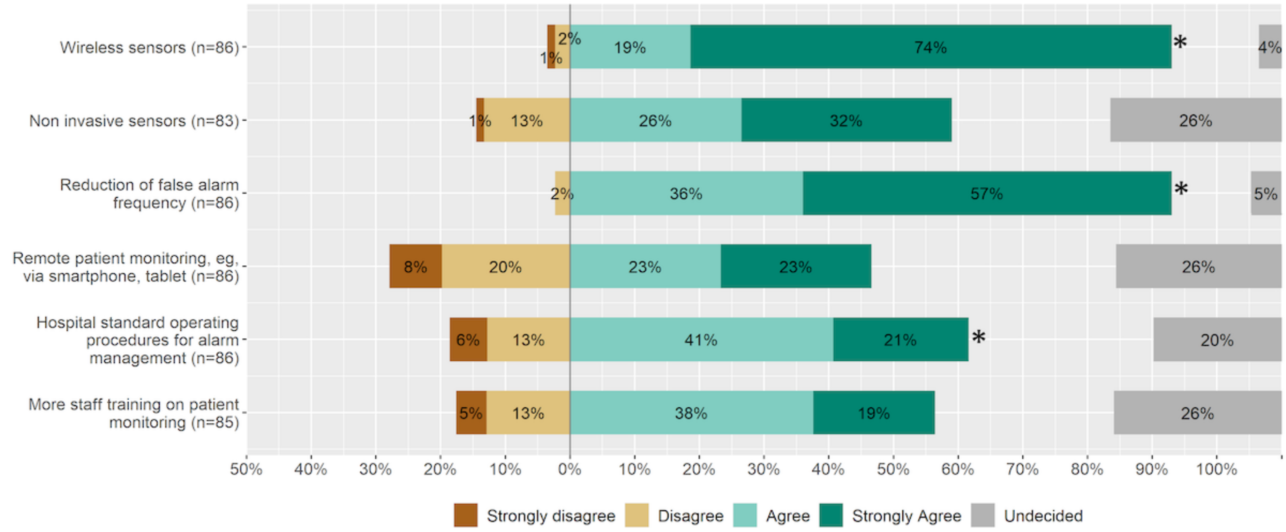


Suggestions for Future Patient Monitoring

Improvements for Future Patient Monitoring

For future patient monitoring, almost all of the ICU staff surveyed requested wireless sensors (n=80, 93% chose “Strongly agree” or “Agree”) and a reduction in false-positive alarms (n=80, 93% chose “Strongly agree” or “Agree”). False-positive alarms may occur due to measurement errors, artifacts, or incorrect settings (Figure 3). Furthermore, respondents wanted a hospital standard operating procedure (SOP) for alarm management (n=53, 62% chose “Strongly agree” or “Agree”). The median responses for the items “Noninvasive sensors,” “Remote patient monitoring,” and “More staff training on patient monitoring” did not significantly differ from the option “Undecided.”

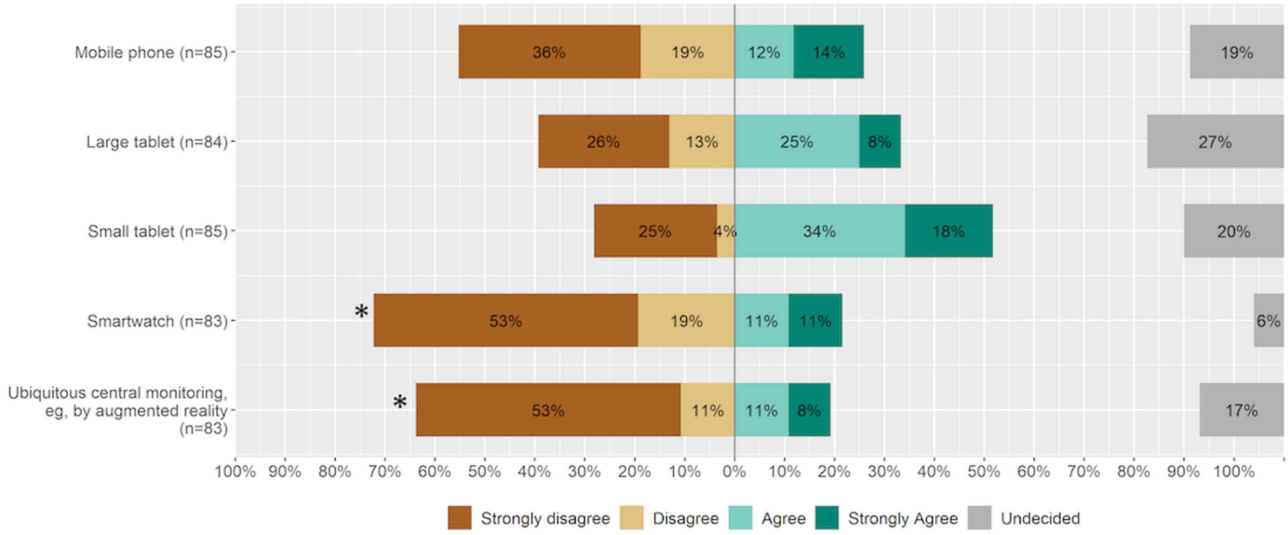
Figure 3. Improvements for future patient monitoring in the ICU. An asterisk indicates statistical significance. ICU: intensive care unit.



Display Devices and Use Cases for Remote Patient Monitoring

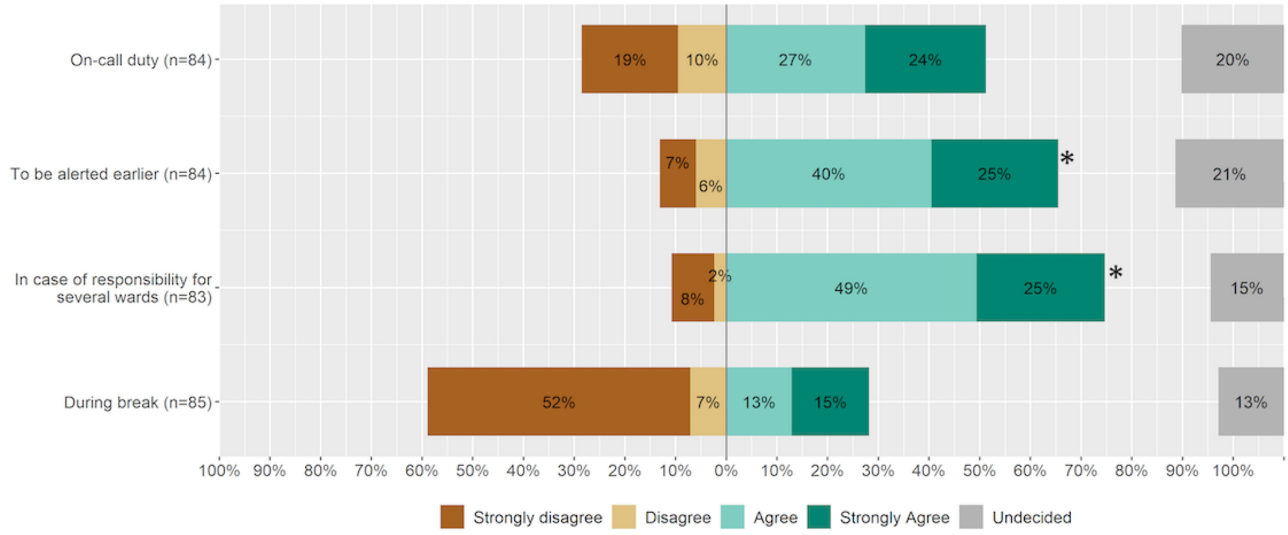
According to the survey results, none of the proposed display devices were desired by ICU staff (Figure 4). The use of smartwatches or augmented reality (AR) glasses in the ICU was rejected by 72% (n=60) and 64% (n=53) of respondents, respectively (those who chose "Strongly disagree" or "Disagree"). With regard to the use of mobile phones for remote patient monitoring, nurses strongly rejected it, while physicians had a neutral attitude toward it.

Figure 4. Suggestions for remote patient monitoring display devices in intensive care medicine for usage on hospital premises. An asterisk indicates statistical significance.



The majority of respondents would appreciate a remote patient monitoring system in an intensive care setting in case they wanted to be alerted earlier (n=55, 65% indicated "Strongly agree" or "Agree") or were responsible for multiple wards (n=62, 74% chose "Strongly agree" or "Agree"; Figure 5). Although not statistically significant, most respondents preferred a remote patient monitoring device for on-call duty, but did not find it useful while taking breaks.

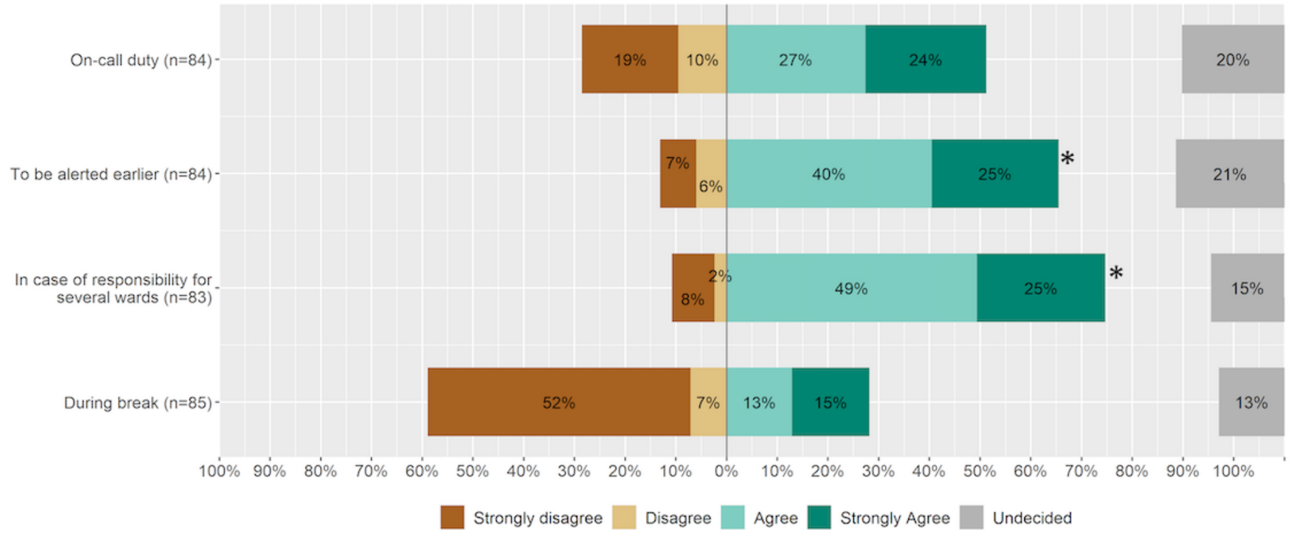
Figure 5. Use cases for remote patient monitoring on hospital premises for intensive care medicine. An asterisk indicates statistical significance.



CDSS

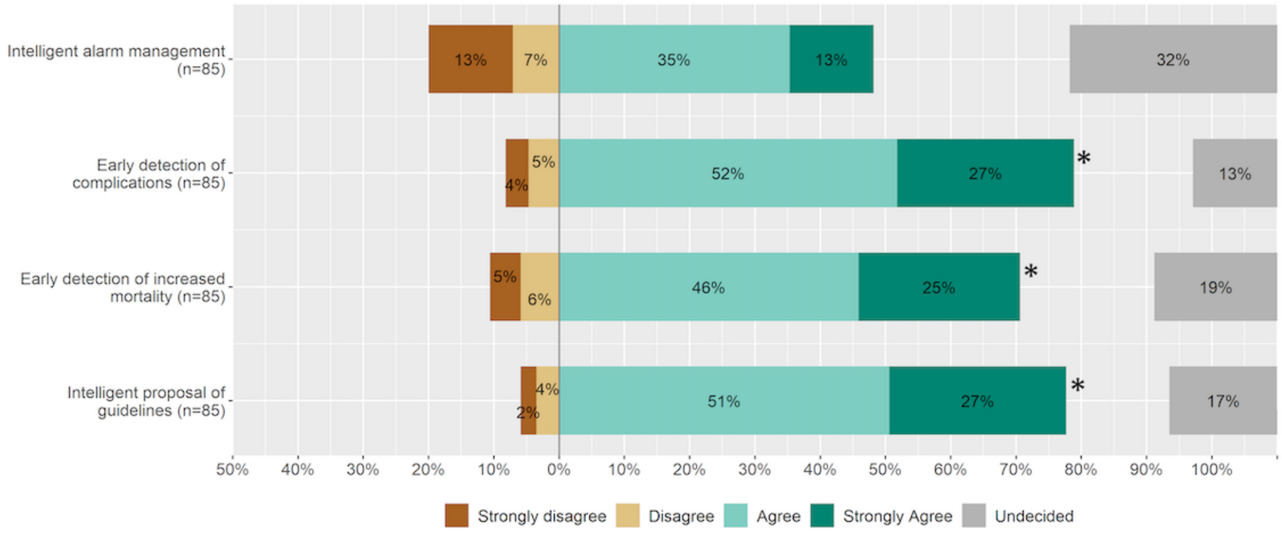
In the future, survey respondents would use a CDSS in the ICU that predicts complications (n=67, 79% chose "Strongly agree" or "Agree") or the risk of mortality of patients (n=60, 71% indicated "Strongly agree" or "Agree") as that intelligently proposes guidelines for therapy and diagnostics (n=66, 78% chose "Strongly agree" or "Agree"; Figure 6). Respondents were inclined to use it for alarm management. Physicians had fewer reservations in using a CDSS with intelligent alarm management than nurses.

Figure 6. Use cases for clinical decision support systems based on artificial intelligence in the ICU. An asterisk indicates statistical significance. ICU: intensive care unit.



Among the factors that users found essential for the use of CDSS, high interoperability (n=79, 93% chose "Strongly agree" or "Agree") and high usability (n=78, 93% indicated "Strongly agree" or "Agree") were deemed most essential. These were followed by the offer of regular staff training with the technology (n=75, 90% chose "Strongly agree" or "Agree") and high transparency of the system (n=66, 78% indicated "Strongly agree" or "Agree"; Figure 7). Most physicians and nurses agreed that regular support (eg, training and workshops) promotes the use of CDSS; more physicians chose "Strongly agree," while more nurses chose "Agree."

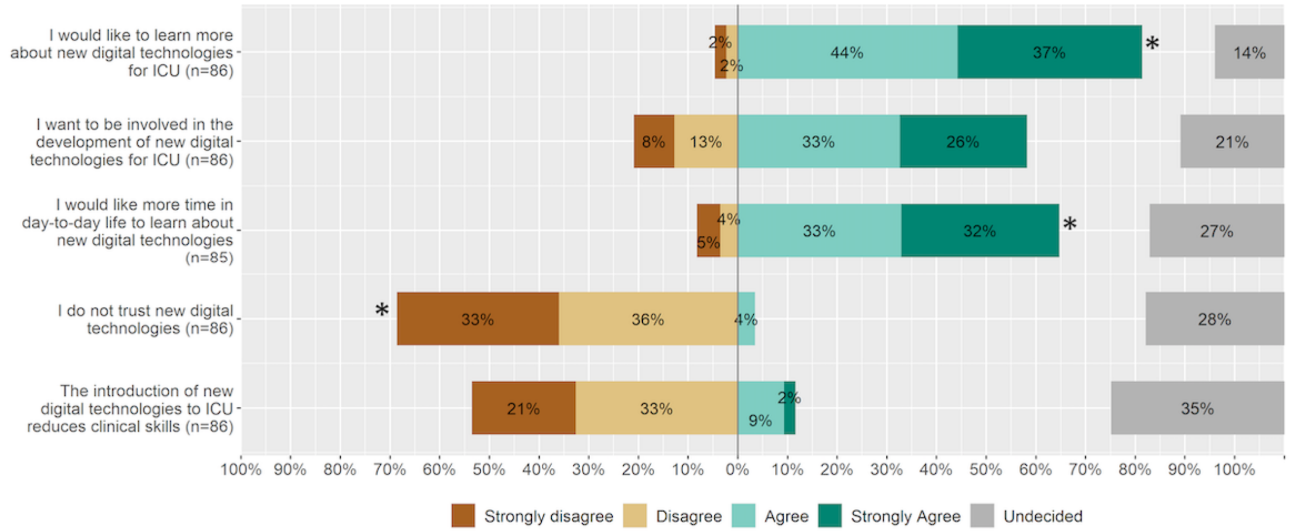
Figure 7. Aspects that promote the usage of clinical decision support systems based on artificial intelligence in the ICU. An asterisk indicates statistical significance. ICU: intensive care unit.



Attitude Toward Novel Technology

Overall, survey respondents were open-minded toward novel technology; among the respondents, 81% (n=70) wanted to know more and 65% (n=55) needed more time to learn about it ("Strongly agree" or "Agree"; Figure 8). The majority (n=59, 69%) disagreed or strongly disagreed on the item "I do not trust new digital technology." Although not statistically significant, 50 respondents (59%) wanted to be involved in the product development of novel digital technologies.

Figure 8. Attitude of ICU staff towards novel digital technology. An asterisk indicates statistical significance. ICU: intensive care unit.



Discussion

Principal Findings

This survey study of ICU staff provides a substantial understanding of the needs and expectations of patient monitoring systems in intensive care medicine from the user's perspective (Textbox 1). Although respondents were confident in using the current patient monitoring system, the high rate of false alarms and the numerous sensor cables were found to potentially interrupt patient care. ICU staff demanded wireless sensors, fewer false alarms, and a hospital SOP for alarm management. Notably, the median replies on display devices for remote patient monitoring did not differ significantly from the option "Undecided," except for the items "Smartwatch" and "Ubiquitous monitoring, eg, through AR," which were both declined. Remote patient monitoring was classified useful for earlier alerts or when responsible for several ICUs. Respondents would use a CDSS based on AI to predict complications, detect increased risk of mortality, and propose guidelines. High transparency, high interoperability, high usability, and regular staff training were all aspects that would promote its usage. Regarding digital literacy, ICU staff was eager to learn more about digital technology and spend more time with it.

Textbox 1. The five most anticipated improvements for patient monitoring by intensive care unit staff.

- Reduction of false alarms
- Implementation of hospital alarm standard operating procedures
- Introduction of wireless sensors
- Introduction of a clinical decision support system based on artificial intelligence
- Enhancement of staff members' digital literacy

Lessons Learned From Today's Patient Monitoring

Notably, we have not observed a proactive call to pioneer new technologies and integrate their respective digital gadgets (eg, smartwatch and AR) into clinical care. Rather, ICU staff looked forward to improvements in the functionality of existing technologies. In line with previous publications, respondents reported that the high rate of false alarms interrupted patient care and demanded a hospital SOP for alarm management [21]. In several studies, implementation of such an alarm management SOP reduced the alarm rate significantly [21,22]. Further temporal analysis of the alarm frequencies per sensor as previously described [23] may find causes for the high rate of false alarms.

It has been reported that cable entanglement is a problem in not only ICUs, but also other places where patients are monitored, such as in operating rooms [24]. Wireless sensors for monitoring vital signs have been tested and implemented several times on stepdown units [6,25]. In many cases, technical requirement analysis (eg, Bluetooth connectivity and interference with other medical devices) was conducted more than a decade ago [26,27]. However, implementation into intensive care routines is still in its infancy [28]. Reasons for this may be the costs associated with developing novel wireless sensors for a high-reliability environment such as the ICU, and technical challenges associated with the need to recharge sensors regularly. In the meantime, cord wraps may facilitate patient transfer with patient monitoring [29].

Remote Patient Monitoring in Intensive Care Medicine

Remote patient monitoring enables clinicians to collect health data via vital sign sensors from patients at location A and electronically transfer this information to location B, where specialists access the data and give health care providers at location A recommendations for managing their patients [4]. Although this is well established in the outpatient sector between the patient's home and the physician [30], the question remains whether this can be supportive to working conditions and patient care in the ICU without a telemedicine context.

Contrary to our preceding qualitative study results, opinions regarding the need for remote patient monitoring in the ICU were divided [8]. There are several industry providers that allow ICU patients to be monitored remotely from anywhere on the premises of the hospital [31-33]. However, scientific evidence of the utility of these devices (eg, for increasing patient safety) seems to be missing. For now, we can summarize that the advantages of on-premise remote patient monitoring for intensive care medicine have to be further quantified by measures such as the reduction of alarms, and improved patient outcomes such as a reduction in patient length of stay.

CDSS in Intensive Care Medicine

As the amount of data as well as the complexity of diseases and treatment of ICU patients are increasing, it seems reasonable to augment the abilities of ICU staff by implementing CDSS based on AI in the ICU. Our results indicate that most of the topics proposed (eg, prediction of mortality, prediction of complications, or proposal of guidelines) were seen as potential use cases for CDSS by ICU staff. For these and several other instances, algorithms already exist that could be adjusted for real-time data [34].

On the path toward implementing CDSS based on AI in intensive care medicine, several barriers have to be overcome [35]. With the introduction of the electronic health record and PDMS in the ICU, the first step has been taken to establish the technical infrastructure, but these systems need to be optimized in interoperability and data quality to act as the basis for complex machine learning processes. To utilize the power of AI as soon as possible, hospital providers should focus on developing data science departments, and introduce standards in implementing novel CDSS tools to rapidly address technical, legal, ethical, and privacy issues.

Transdisciplinary Research and Development

Clinical teams in ICUs are used to working closely together in multidisciplinary teams. This could be advantageous when adding further professions to the team for transdisciplinary research and the development of medical devices for intensive care medicine [36]. Our survey results show that ICU staff members are open to learning more about technology and are even willing to support product development in some cases. Thus, a clinical data scientist with formal medical training could be part of the ICU team as well as the product development team alongside engineers from a medical manufacturer [22,37]. This transdisciplinary approach should be piloted in further studies, to assess the effects on mutual exchange and innovation potential.

As much as the transdisciplinary approach is supported, blunt confidence in user feedback will mainly improve existing devices, as our study prominently indicates, which does not necessarily foster the discovery of disruptive technologies [38], such as avatar-based patient monitoring [39,40] or smart glasses [41]. More than cooperation, transdisciplinarity refers to the development of common theories, mutual observation, and search for challenges and needs [42]. Hackathons (weekend innovation events) are an excellent playground for transdisciplinary work, and participation should be encouraged and remunerated by medical manufacturers and hospital providers [43].

Limitations

With this survey study among ICU staff, we identified the most anticipated improvements for patient monitoring in the ICU from the user perspective. However, several limitations apply to this study. It is important to note that the developed questionnaire did not include questions of established reliability or validity; the data were collected at a single hospital in Germany; the number of participating physicians was small, making statements about group comparisons susceptible to coincidence; and the response rate was moderate. Due to the online collection of data, the participation of ICU staff with less technical affinity may have been reduced. Further studies including a sample size calculation and randomized sample collection would reduce the risk of bias.

Whether the findings (eg, introducing wireless patient monitoring sensors) actually lead to an improvement in working conditions and patients’ quality of life or quality of care in the ICU can only be ascertained by further studies. Finally, a bias due to the deployment of the Vital Sync virtual patient monitoring platform in 1 of the 4 ICUs cannot be ruled out with certainty.

Conclusion

This survey study among ICU staff revealed anticipated key improvements for patient monitoring in intensive care medicine from the user perspective. We did not observe a proactive call to pioneer new technologies and integrate their respective digital gadgets (eg, smartwatch and AR) into clinical routine. Instead, ICU staff looked forward to improvements in the functionality of existing technologies. Particularly, hospital providers and medical device manufacturers should focus on reducing false alarms, implementing hospital alarm SOPs, introducing wireless sensors, preparing for CDSS based on AI, and enhancing the digital literacy of ICU staff. In the medium term, our results may contribute to the user-centered transfer of digital technologies into practice to alleviate challenges in intensive care medicine, such as those recently caused by COVID-19.

Acknowledgments

We express our gratitude to the ICU staff for their participation in this study. We would further like to thank the German Research Foundation (DFG), the Open Access Publication Funds of the Charité – Universitätsmedizin Berlin, and the Einstein Center Digital Future as well as the Medtronic for providing the aforementioned devices free of charge.

Authors' Contributions

CS had the idea for shared decision allocation and initiated the implementation of remote patient monitoring in the intensive care unit. The study was conceived by ASP, CS, and FB. ASP conducted data acquisition, supported by LM and FS. ASP and MS analyzed the data, supported by HK, who provided expertise in statistics. ASP wrote the manuscript, supported by LM, MS, and FS. HK contributed to the study's design and interpretation of results from a psychologist's point of view. FB supervised all parts of the study. All authors critically reviewed and approved the manuscript.

Conflicts of Interest

CS and FB report funding from Medtronic. The other authors do not have conflicts to declare.

Multimedia Appendix 1
Final questionnaire.
[XLSX File (Microsoft Excel File), 55 KB - jmir_v22i6e19091_app1.xlsx]

Multimedia Appendix 2
Survey item medians and bootstrap CIs.
[XLSX File (Microsoft Excel File), 11 KB - jmir_v22i6e19091_app2.xlsx]

Multimedia Appendix 3
Survey raw data.
[XLSX File (Microsoft Excel File), 22 KB - jmir_v22i6e19091_app3.xlsx]

Multimedia Appendix 4
Distribution of item responses of physicians and nurses.
[PNG File , 3159 KB - jmir_v22i6e19091_app4.png]

References

1. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-06-05]
2. De Georgia M, Loparo K. Neurocritical Care Informatics: Translating Raw Data into Bedside Action Internet. Berlin, Heidelberg: Springer Berlin Heidelberg; 2020.
3. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/nejmoa1901183]
4. Koenig M. Telemedicine in the ICU. Basel, Switzerland: Springer International Publishing; 2019.
5. Michard F, Bellomo R, Taenzer A. The rise of ward monitoring: opportunities and challenges for critical care specialists. Intensive Care Med 2019 May 27;45(5):671-673. [doi: 10.1007/s00134-018-5384-5] [Medline: 30259073]
6. Weenk M, Koeneman M, van de Belt TH, Engelen LJ, van Goor H, Bredie SJ. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation 2019 Mar;136:47-53. [doi: 10.1016/j.resuscitation.2019.01.017] [Medline: 30685546]
7. De Georgia MA, Kaffashi F, Jacono FJ, Loparo KA. Information technology in critical care: review of monitoring and data acquisition systems for patient care and research. The Scientific World Journal 2015;2015:727694-727699 [FREE Full text] [doi: 10.1155/2015/727694] [Medline: 25734185]
8. Poncette A, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, et al. Clinical Requirements of Future Patient Monitoring in the Intensive Care Unit: Qualitative Study. JMIR Med Inform 2019 Apr 30;7(2):e13064 [FREE Full text] [doi: 10.2196/13064] [Medline: 31038467]
9. Burns KEA, Duffett M, Kho ME, Meade MO, Adhikari NKJ, Sinuff T, et al. A guide for the design and conduct of self-administered surveys of clinicians. CMAJ 2008 Jul 29;179(3):245-252 [FREE Full text] [doi: 10.1503/cmaj.080372] [Medline: 18663204]
10. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
11. Wessel D, Attig C, Franke T. ATI-S - An Ultra-Short Scale for Assessing Affinity for Technology Interaction in User Studies. 2019 Presented at: MuC'19: Proceedings of Mensch und Computer 2019; 2019; Hamburg, Germany. [doi: 10.1145/3340764.3340766]
12. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

13. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
14. R Core Team. R Foundation for Statistical Computing. Vienna, Austria: A language and environment for statistical computing; 2018. URL: https://www.R-project.org/ [accessed 2020-06-05]
15. Wickham H. tidyverse: Easily Install and Load the Tidyverse. URL: https://CRAN.R-project.org/package=tidyverse [accessed 2020-03-09]
16. Lüdecke D, Bartel A, Schwemmer C, Powell C, Djalovski A. sjPlot: Data Visualization for Statistics in Social Science. URL: https://CRAN.R-project.org/package=sjPlot [accessed 2020-03-10]
17. Revelle W. 2020. psych: Procedures for Psychological, Psychometric, and Personality Research URL: https://CRAN.R-project.org/package=psych [accessed 2020-03-26]
18. Bray A, Ismay C, Chasnovski E, Baumer B, Cetinkaya-Rundel M, Laderas T, et al. infer: Tidy Statistical Inference Internet. 2019. URL: https://CRAN.R-project.org/package=infer [accessed 2020-03-10]
19. Duhm J, Fleischmann R, Schmidt S, Hupperts H, Brandt SA. Mobile Electronic Medical Records Promote Workflow: Physicians' Perspective From a Survey. JMIR mHealth uHealth 2016 Jun 06;4(2):e70 [FREE Full text] [doi: 10.2196/mhealth.5464] [Medline: 27268720]
20. Haukoos JS, Lewis RJ. Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. Acad Emerg Med 2005 Apr 01;12(4):360-365 [FREE Full text] [doi: 10.1197/j.aem.2004.11.018] [Medline: 15805329]
21. Bach TA, Berglund L, Turk E. Managing alarm systems for quality and safety in the hospital setting. BMJ Open Qual 2018 Jul 25;7(3):e000202 [FREE Full text] [doi: 10.1136/bmjoq-2017-000202] [Medline: 30094341]
22. Pater CM, Sosa TK, Boyer J, Cable R, Egan M, Knilans TK, et al. Time series evaluation of improvement interventions to reduce alarm notifications in a paediatric hospital. BMJ Qual Saf 2020 Jan 20:2019. [doi: 10.1136/bmjqs-2019-010368] [Medline: 31959714]
23. Wilken M, Hüske-Kraus D, Röhrig R. Alarm Fatigue: Using Alarm Data from a Patient Data Monitoring System on an Intensive Care Unit to Improve the Alarm Management. Studies in Health Technology and Informatics 2019;273:281. [doi: 10.29007/5vpj]
24. Tscholl D, Handschin L, Rössler J, Weiss M, Spahn D, Nöthiger C. It's not you, it's the design - common problems with patient monitoring reported by anesthesiologists: a mixed qualitative and quantitative study. BMC Anesthesiology 2019;19(1):a. [doi: 10.21203/rs.2.238/v2]
25. Glasin J, Henricson J, Lindberg L, Wilhelms D. Wireless vitals-Proof of concept for wireless patient monitoring in an emergency department setting. J Biophotonics 2019 Apr 13;12(4):e201800275. [doi: 10.1002/jbio.201800275] [Medline: 30306737]
26. Wallin MKEB, Wajntraub S. Evaluation of Bluetooth as a replacement for cables in intensive care and surgery. Anesth Analg 2004 Mar;98(3):763-767. [doi: 10.1213/01.ane.0000097186.99167.ba] [Medline: 14980934]
27. Paksuniemi M, Sorvoja H, Alasaarela E, Myllyla R. Wireless sensor and data transmission needs and technologies for patient monitoring in the operating room and intensive care unit. 2005 Presented at: 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2005; Shanghai, China p. 5182. [doi: 10.1109/iembs.2005.1615645]
28. Chung HU, Rwei AY, Hourlier-Fargette A, Xu S, Lee K, Dunne EC, et al. Skin-interfaced biosensors for advanced wireless physiological monitoring in neonatal and pediatric intensive-care units. Nat Med 2020 Mar 11;26(3):418-429. [doi: 10.1038/s41591-020-0792-9] [Medline: 32161411]
29. Lambert DH. Cord Wraps Facilitate Patient Transfer. Anesthesia & Analgesia 2016;123(1):257. [doi: 10.1213/ane.0000000000001376]
30. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. The Lancet 2018 Sep;392(10152):1047-1057. [doi: 10.1016/s0140-6736(18)31880-4]
31. Medtronic - Vital Sync Virtual Patient Monitoring Platform. URL: https://www.medtronic.com/covidien/en-us/products/health-informatics-and-monitoring/vital-sync-virtual-patient-monitoring-platform-2-6.html [accessed 2019-01-15]
32. Philips. IntelliVue Mobile Caregiver. URL: https://www.usa.philips.com/healthcare/product/HCNOCTN197/intellivue-mobile-caregiver-mobile-app-for-patient-monitoring-data [accessed 2020-03-17]
33. Draeger. Mobile Patient Watch. URL: https://www.draeger.com/en_sea/Hospital/Products/Medical-Systems-Solutions/Clinical-Applications/Mobile-Patient-Watch [accessed 2020-03-17]
34. Mlodzinski E, Stone DJ, Celi LA. Machine Learning for Pulmonary and Critical Care Medicine: A Narrative Review. Pulm Ther 2020 Jun 5;6(1):67-77 [FREE Full text] [doi: 10.1007/s41030-020-00110-z] [Medline: 32048244]
35. Komorowski M. Artificial intelligence in intensive care: are we there yet? Intensive Care Med 2019 Sep 24;45(9):1298-1300. [doi: 10.1007/s00134-019-05662-6] [Medline: 31236638]
36. Özcan E, Birdja D, Simonse L, Struijs A. Alarm in the ICU! Envisioning Patient MonitoringAlarm Management in Future Intensive Care Units. Basel, Switzerland: Springer International Publishing; 2019.

37. Hüske-Kraus D, Wilken M, Röhrig R. Measuring Alarm System Quality in Intensive Care Units. Zukunft der Pflege Tagungsband der 1. Clusterkonferenz 2018 2018:89-93. [doi: 10.29007/5vpj]
38. Cohen AB, Martin SS. Innovation without integration. NPJ Digit Med 2020 Feb 3;3(1):15 [FREE Full text] [doi: 10.1038/s41746-020-0220-z] [Medline: 32025575]
39. Garot O, Rössler J, Pfarr J, Ganter MT, Spahn DR, Nöthiger CB, et al. Avatar-based versus conventional vital sign display in a central monitor for monitoring multiple patients: a multicenter computer-based laboratory study. BMC Med Inform Decis Mak 2020 Feb 10;20(1):26 [FREE Full text] [doi: 10.1186/s12911-020-1032-4] [Medline: 32041584]
40. Tscholl DW, Rössler J, Handschin L, Seifert B, Spahn DR, Nöthiger CB. The Mechanisms Responsible for Improved Information Transfer in Avatar-Based Patient Monitoring: Multicenter Comparative Eye-Tracking Study. J Med Internet Res 2020 Mar 16;22(3):e15070 [FREE Full text] [doi: 10.2196/15070] [Medline: 32175913]
41. Romare C, Skär L. Smart Glasses for Caring Situations in Complex Care Environments: Scoping Review. JMIR mHealth uHealth 2020 Apr 20;8(4):e16055 [FREE Full text] [doi: 10.2196/16055] [Medline: 32310144]
42. Méndez F. Transdiscipline and research in health: science, society and decision making. Colomb Med (Cali) 2015 Sep 30;46(3):128-134 [FREE Full text] [Medline: 26600628]
43. Poncette A, Rojas P, Hofferbert J, Valera Sosa A, Balzer F, Braune K. Hackathons as Stepping Stones in Health Care Innovation: Case Study With Systematic Recommendations. J Med Internet Res 2020 Mar 24;22(3):e17004 [FREE Full text] [doi: 10.2196/17004] [Medline: 32207691]

Abbreviations

AI: artificial intelligence
AR: augmented reality
ATI: Affinity for Technological Interaction
CDSS: clinical decision support system
COVID-19: coronavirus disease
ICU: intensive care unit
PDMS: patient data management system
QR: quick response
SOP: standard operating procedure

Edited by G Eysenbach; submitted 13.04.20; peer-reviewed by MA Bahrami, J Rössler; comments to author 30.04.20; revised version received 05.05.20; accepted 13.05.20; published 19.06.20.

Please cite as:
Poncette AS, Mosch L, Spies C, Schmieding M, Schiefenhövel F, Krampe H, Balzer F
Improvements in Patient Monitoring in the Intensive Care Unit: Survey Study
J Med Internet Res 2020;22(6):e19091
URL: http://www.jmir.org/2020/6/e19091/
doi:10.2196/19091
PMID:32459655

Original Paper

Active Surveillance of Adverse Events Following Human Papillomavirus Vaccination: Feasibility Pilot Study Based on the Regional Health Care Information Platform in the City of Ningbo, China

Zhike Liu[1*], MD; Liang Zhang[2*], BSc; Yu Yang[3], MD; Ruogu Meng[3], MD; Ting Fang[2], MM; Ying Dong[2], MM; Ning Li[2], MM; Guozhang Xu[2], MM; Siyan Zhan[1], MD

[1]Department of Epidemiology and Biostatistics, Peking University Health Science Center, Beijing, China
[2]Ningbo Municipal Center for Disease Control and Prevention, Ningbo, China
[3]National Institute of Health Data Science, Peking University, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Siyan Zhan, MD
Department of Epidemiology and Biostatistics
Peking University Health Science Center
38 Xueyuan Road
Haidian District
Beijing, 100191
China
Phone: 86 13641084052
Email: siyan-zhan@bjmu.edu.cn

Abstract

Background: Comprehensive safety data for vaccines from post-licensure surveillance, especially active surveillance, could guide administrations and individuals to make reasonable decisions on vaccination. Therefore, we designed a pilot study to assess the capability of a regional health care information platform to actively monitor the safety of a newly licensed vaccine.

Objective: This study aimed to conduct active surveillance of human papillomavirus (HPV) vaccine safety based on this information platform.

Methods: In 2017, one of China's most mature information platforms with superior data linkage was selected. A structured questionnaire and open-ended interview guidelines were developed to investigate the feasibility of active surveillance following HPV vaccination using the regional health care information platform in Ningbo. The questionnaire was sent to participants via email, and a face-to-face interview was conducted to confirm details or resolve discrepancies.

Results: Five databases that could be considered essential to active surveillance of vaccine safety were integrated into the platform starting in 2015. Except for residents' health records, which had a coverage rate of 87%, the data sources covered more than 95% of the records that were documented in Ningbo. All the data could be inherently linked using the national identity card. There were 19,328 women who received the HPV vaccine, and 37,988 doses were administered in 2017 and 2018. Women aged 30-40 years accounted for the largest proportion. Quadrivalent vaccination accounted for 73.1% of total vaccination, a much higher proportion than that of bivalent vaccination. Of the first doses, 60 (60/19,328, 0.31%) occurred outside Ningbo. There were no missing data for vaccination-relevant variables, such as identity card, vaccine name, vaccination doses, vaccination date, and manufacturer. ICD-10 coding could be used to identify 9,180 cases using a predefined list of the outcomes of interest, and 1.88% of these cases were missing the identity card. During the 90 days following HPV vaccination, 4 incident cases were found through the linked vaccination history and electronic medical records. The combined incident rate of rheumatoid arthritis, optic neuritis, and Henoch-Schonlein purpura was 8.84/100,000 doses of bivalent HPV, and the incidence rate of rheumatoid arthritis was 3.75/100,000 doses of quadrivalent HPV.

Conclusions: This study presents an available approach to initiate an active surveillance system for adverse events following HPV vaccination, based on a regional health care information platform in China. An extended observation period or the inclusion of additional functional sites is warranted to conduct future hypothesis-generating and hypothesis-confirming studies for vaccine safety concerns.

(J Med Internet Res 2020;22(6):e17446) doi:10.2196/17446

KEYWORDS

safety; HPV; human papillomavirus; vaccine; active surveillance

Introduction

Vaccines, unlike drugs, are generally administered to healthy populations, so the target sample is much larger, especially for birth cohorts, children, or pregnant women. Concerns around vaccine safety, especially for rare adverse events, are increasing with the success of vaccines for the control of vaccine-preventable diseases [1,2]. These concerns can cause a lack of confidence in vaccines and further vaccine hesitancy, defined as a delay in the acceptance or refusal of vaccines despite the availability of vaccination services, that impede or undermine the efforts of an immunization program [3,4]. Although pre-licensure trials can assess the relationship between a vaccine and common adverse events, this type of trial does not have sufficient power to determine the general risk of rare diseases. Therefore, comprehensive safety data from post-licensure surveillance should be used to evaluate the benefit-risk ratio of vaccines, allowing evidence-based decisions by individuals and health organizations [5,6].

Surveillance of adverse events following immunization (AEFI), especially active surveillance for AEFI, should be an essential part of an immunization program to guarantee the safety of vaccines and help establish public confidence [7]. Worldwide, there are a variety of approaches for active AEFI surveillance. Of these approaches, two are the most recognized and replicated due to their robustness and continuity [7]: Vaccine Safety Datalink (VSD), consisting of population-based distributed datasets [8,9], and IMPACT, consisting of targeted hospital-based surveillance [10,11]. These systems were primarily created by developed countries, and limited such systems are available in developing countries [12]. Because of its large population, vast area, and different racial and cultural backgrounds, evidence of vaccine safety in China, especially from active surveillance, can contribute to global vaccine safety efforts.

In recent years, the number of electronic health data sources and interest in their applications have rapidly increased with the advances in medical information systems in China [13], and these electronic health databases could provide the potential for drug safety surveillance and pharmacoepidemiology [14]. Among the electronic health data sources, regional health data sources have been successfully used for drug safety evaluations and surveillance of chronic diseases [15,16]. However, very little is known about the capacity of this data source for active surveillance of adverse events following immunization, and increasing importance has been given to feasibility assessments as essential technical evaluations before starting any vaccine epidemiologic study. Here, we aimed to conduct a feasibility assessment for active surveillance of human papillomavirus (HPV) vaccine safety using the regional health data platform in the city of Ningbo, including a description of the database stakeholders and components, evaluation of the completeness of data relevant to HPV vaccine safety, and an examination of the linkage of data between different data sources.

Methods

Study Setting and Data Sources

Ningbo is a city in the Zhejiang province of China (28°51’ – 30°33’ N, 120°55’ – 122°16’ E). It is located in the south wing of the Yangtze River Delta and consists of 4 counties (Yuyao, Cixi, Ninghai, and Xiangshan) and 6 districts (Yinzhou, Haishu, Jiangbei, Zhenhai, Beilun, and Fenghua), with a total land area of 9365 km^2. In 2017, the city population was >8 million people, and nearly 6 million are part of the hukou (ie, registration of an individual) of Ningbo. The resident population is relatively stable, with a 4.1% out-migration rate and 11.1% in-migration rate.

The regional health care information platform in Ningbo was initially devised and launched in 2011 by the Health Commission of Ningbo to construct a united and standardized medical information network and facilitate health care services. By 2015, the database covered almost all health-related activities of residents within this area, and more than 87% of the residents were registered in the database with a valid health care identifier. In 2016, the database was approved and awarded a Four Grade Class A, which at the time was the top achievement level by the Standardization and Maturity Measurement of Regional Health Information Interconnection by the National Health Commission of China.

Questionnaire Design and Field Investigation

Based on information in previous studies, we developed the questionnaire and open-ended interview guidelines to evaluate the feasibility of active surveillance following HPV vaccination using the regional health care information platform in Ningbo. To finalize the questionnaire and guidelines, four experts from pharmacoepidemiology, vaccine safety, medical informatics, and database administration independently reviewed the documents and provided feedback. Then, a 2-step survey was conducted to collect database information from the data owners, management, and other staff. First, we sent the questionnaire, its instructions, a comprehensive description of the study purpose, and the completion deadline to the database staff via email. After receiving the questionnaire responses, we confirmed any potential discrepancies with Ningbo through telephone communication. Second, we conducted face-to-face interviews, using the open-ended interview guidelines, with relevant staff during field visits.

The questionnaire about the database consisted of 7 parts: (1) basic information in the database; (2) information in electronic medical records (EMR) in the database; (3) vaccination registration and information about maternal and child health care in the database; (4) method and possibility of linking data between different data sources; (5) age distribution of HPV vaccine–protected women in 2017; (6) completeness of core variables relevant to active surveillance of vaccine safety in 2017; and (7) crucial outcome information in EMR in 2017.

The interview guideline was mainly composed of 3 parts: (1) general characteristics of the database; (2) general characteristics of the essential information system for active surveillance, including vaccination registration, EMR, maternal and child health care information, cancer registration, death registration; and (3) other several specific questions.

The study was approved by the Peking University Institutional Review Board (IRB 00001052-18016). Moreover, all the tasks were performed by staff of the data owner, and investigators could not access the raw data throughout the study period.

Results

Stakeholders and Database Components in Ningbo

The Ningbo database was constructed and is owned by the Ningbo Health Commission (NHC), and the Ningbo Centers for Disease Control and Prevention (CDC) can operate and develop this database with permission from the NHC. There are two primary data sources in the database, including the digital CDC platform and digital hospital platform. The CDC controls the former, which mainly focuses on public health and provides a unified information system for the whole city, including integrated primary health care, chronic or infectious disease surveillance, vaccine registration (children and adults), and death registration. Data in the hospital digital platform are collected during regular clinical services provided in hospitals, such as from EMR, and the information system usually differs by hospital, but the NHC has developed harmonized standards to transform and upload these data, which are in a nearly structured format in the Ningbo database. Moreover, the NHC and CDC have endeavored to improve the data quality in the database through official regulations, upgrading of standards, and regular quality checks. The local CDC has also established an institutional review board to review ethical issues and guarantee privacy protection during the process of scientific research.

Characteristics of the Data Sources for Active Surveillance of HPV Vaccine Safety

Table 1 summarizes the characteristics of the five data sources related to active surveillance for HPV vaccine safety in the Ningbo database; data from these sources were updated and uploaded into the database daily. First, residents’ electronic health records were created primarily by general physicians when the residents visited medical and health institutions for services; these records currently cover >87% of the permanent residents in Ningbo. Second, all vaccination records, including free vaccines provided by the Expanded Programme on Immunization or vaccines paid for out-of-pocket, for infants and children aged <6 years are registered in the immunization program system. It was a compulsory requirement from the Ningbo CDC beginning in May 2017 that vaccination records for adults be registered. Third, starting in 2015, the platform integrated new and historical EMR data from all public hospitals and most of the large private hospitals in the city. Fourth, the maternal and child health care information system includes data recorded during health care services provided for prenatal examinations, maternal deliveries, and newborn follow-ups (until the infant was 1 year old), which resulted in an increase in the local pregnancy and maternal registration rate to 97% by 2016. Finally, the death surveillance information system primarily includes identity information and medical certificates of the dead and can trace death records outside of the city of Ningbo but inside Zhejiang province.

Table 1. Characteristics of the data sources related to HPV vaccine safety surveillance.

Data and population	Year surveillance began	Coverage	Main purpose
RHR[a]			
Residents	2010	>87% of permanent residents	Basic information, health checks, special population primary care
IP[b]			
Children	2005	100% for 167 vaccination clinics	Basic information, vaccination records, vaccine cold chain
Adults	2017		
EMR[c]			
Outpatients[d]	2015	All 221 public health hospitals and 28 large private hospitals	Diagnoses, prescriptions, medical examinations, lab tests
Inpatients	2015		
MCH[e]			
Maternal patients	2015	> 97% maternal registration rate	Prenatal examinations, birth deliveries, newborn follow-ups
Newborns	2015		
DS[f]			
Residents	2010	Permanent residents	Identity, death certificates

[a] RHR: residents' health records.

[b] IP: immunization program.

[c] EMR: electronic medical record.

[d] includes outpatient and emergency visits.

[e] MCH: maternal and child health care.

[f] DS: death surveillance.

Characteristics of HPV Vaccination in the Immunization Program

From 2017 to 2018, there were two types of HPV vaccines: bivalent and quadrivalent. Both were only available for women in this district. Table 2 describes the characteristics of the population who received the HPV vaccination in Ningbo. During this period, 19,328 women received at least one dose of the HPV vaccine. Women aged 30-40 years represented the largest proportion. Quadrivalent vaccination was administered more often than bivalent vaccination. Moreover, 60 (60/19,328, 0.31%) first doses seemingly occurred outside Ningbo, and these data were not traced back to the database.

The completeness of critical variables relevant to HPV vaccination (eg, recipient identity card [ID], name, gender, birth date, address, vaccine name, vaccination date, vaccination manufacturer, batch number) was also investigated, and for these variables, there were no missing data. The system did not have information for education, occupation, height, weight, smoking habits, and alcohol consumption.

Table 2. Characteristics of the women who underwent HPV vaccination.

Variables	Total (n=19,328), n (%)	First dose (n=19,268), n (%)	Second dose (n=13,404), n (%)	Third dose (n=5316), n (%)
Age group (years)				
9-15	1541 (7.97)	1532 (7.95)	1343 (10.02)	613 (11.53)
16-20	959 (4.96)	943 (4.89)	806 (6.01)	491 (9.24)
21-25	2735 (14.15)	2721 (14.12)	1961 (14.63)	1064 (20.02)
26-30	3221 (16.66)	3215 (16.69)	1966 (14.67)	634 (11.93)
31-35	4146 (21.45)	4138 (21.48)	2554 (19.05)	759 (14.28)
36-40	4117 (21.30)	4112 (21.34)	2753 (20.54)	902 (16.97)
41-45	2609 (13.50)	2607 (13.53)	2021 (15.08)	853 (16.05)
Vaccine				
Bivalent	5194 (26.87)	5147 (26.71)	4445 (33.16)	1721 (32.37)
Quadrivalent	14134 (73.13)	14121 (73.29)	8959 (66.84)	3595 (67.63)

Identification of the Outcome of Interest in EMR

Table 3 shows the ability of the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes to identify the safety outcomes of interest in EMR as well as the incident cases of diseases 90 days after HPV vaccination. All visit records of the cases were retrieved from the EMR using standard query language and the ICD-10 code, and duplicate records were removed. As a result, for data from 2017 and 2018, ICD-10 codes can be used to identify all the outcomes of interest in women aged 9-45 years. Among these diseases, there were 1867 cases of systemic lupus erythematosus, which was the largest proportion of cases. In contrast, there were only 2 cases of acute disseminated encephalomyelitis, which was the smallest proportion of cases. Except for that of Henoch-Schonlein purpura (HSP; 118/1606, 7.35%), the rate of missing IDs for other diseases was near or <5%, and the total rate of missing IDs was 1.88% (173/9180). Moreover, we monitored the risk of specific adverse events during the 90 days after HPV vaccination. For all the doses, we found 4 incident cases with no history of the disease at least 1 year before HPV vaccination. Two new cases, one each of rheumatoid arthritis (RA) and optic neuritis (ON), were diagnosed within 30 days of bivalent HPV vaccination, and one new case of HSP was diagnosed within 90 days of bivalent HPV vaccination. One new case of RA emerged within 90 days after quadrivalent HPV vaccination.

Hospitals not included in the database could not diagnose nor treat the outcomes of interest. Regarding case validation, structured data from imaging examinations, laboratory tests, histological examinations, and other medical examinations could be found directly in the database, while nonstructured or free-text information, such as the chief complaint, may have needed to be retrieved from the hospital.

Table 3. Outcomes among women aged 9-45 years recorded in electronic medical records.

Disease	ICD-10[a]	n[b]	ID[c] missing, n (%)	New cases 30 days post-HPV[d] vaccination, n	New cases 90 days post-HPV vaccination, n
Rheumatologic/autoimmune diseases					
SLE[e]	M32	1867	39 (2.09)	0	0
RA[f]	M05, M06	2878	72 (2.50)	1	2
JRA[g]	M08.0	52	3 (5.77)	0	0
Inflammatory bowel disease (IBD)					
Crohn disease	K50	135	6 (4.44)	0	0
Ulcerative colitis	K51	787	17 (2.16)	0	0
Autoimmune endocrine conditions					
Type 1 diabetes	E10	369	18 (4.88)	0	0
Autoimmune thyroiditis	E06.3	126	2 (1.59)	0	0
Graves’ disease	E05.0	362	5 (1.38)	0	0
Autoimmune neurologic and ophthalmic conditions					
Multiple sclerosis	G35	30	0 (0)	0	0
ADEM[h]	G04.0	2	0 (0)	0	0
GBS[i]	G61.0	11	0 (0)	0	0
Neuromyelitis optica	G36.0	35	1 (2.86)	0	0
ON[j]	H46	346	16 (4.62)	1	1
Others					
ITP[k]	D69.3|4	279	4 (1.43)	0	0
HSP[l]	D69.0	1606	118 (7.35)	0	1
Bell’s palsy	G51.0	105	5 (4.76)	0	0
VTE[m]	I82.8|9	59	3 (5.08)	0	0
Raynaud’s disease	I73.0	131	1 (0.76)	0	0
Total	N/A[n]	9180	173 (1.88)	2	4

[a] ICD-10: 10th revision of the International Statistical Classification of Diseases and Related Health Problems.

[b] duplicate records were removed, and the number of patients was calculated using an encoded identifier consisting of the identity card or name, gender, birth date, and address.

[c] ID: identity card, which is unique for each individual in China.

[d] HPV: human papillomavirus.

[e] SLE: systemic lupus erythematosus.

[f] RA: rheumatoid arthritis.

[g] JRA: juvenile rheumatoid arthritis.

[h] ADEM: acute disseminated encephalomyelitis.

[i] GBS: Guillain-Barre syndrome.

[j] ON: optic neuritis.

[k] ITP: primary immune thrombocytopenia.

[l] HSP: Henoch-Schonlein purpura.

[m] VTE: venous thromboembolism.

[n] N/A: not applicable.

Discussion

Principal Findings

A feasibility assessment is the first step to improve the quality of study protocols and accelerate regulatory approvals and, in turn, the start of an actual study [17]. The regional health care information platform in Ningbo covers the entire city and has integrated the essential data sources for active surveillance of the HPV vaccine, including immunization program registration, EMR, resident health records, and death surveillance; these data can be linked using the unique national ID or encoded identifier. More importantly, this database can sensitively identify the outcome of interest through ICD-10 coding, so we can actively monitor risks following HPV vaccination by linking HPV vaccination records with adverse events from EMR, death, or other data sources. Therefore, this database could be available and feasible in research for active surveillance of adverse events following HPV vaccination.

Comparison With Prior Work

The Chinese CDC established the AEFI Surveillance System, which is a passive national surveillance system covering more than 29 provinces, in 2015 [18]. However, regarding active surveillance, the relevant studies have used traditional methods for timely collection of adverse events, such as by telephone or daily cards, that were designed and performed during the H1N1 pandemic [19]. Plus, we have not found a study that attempted to construct a sustained active surveillance system for AEFI despite the large population and vast geographical area in China [7]. Internationally, some countries have developed sustainable active surveillance systems for adverse events using population-based databases that link vaccine history with outcomes, such as the VSD [9] and Post-Licensure Rapid Immunization Safety Monitoring program [20]. In this study, considering that the crucial data components in the Ningbo database are similar to those in the VSD and Post-Licensure Rapid Immunization Safety Monitoring program, we consider this to be pioneering work in China – to construct a continuous method for actively monitoring AEFI during the post-marketing phase of vaccines. Moreover, we can also link vaccination history with death records in the Ningbo database, which might reduce potential bias from right-censored data.

In this study, of the specific list of outcomes, we identified 4 incident cases within the 90 days following HPV vaccination. The combined incident rate of RA, ON, and HSP was 8.84/100,000 doses of bivalent HPV, and the incident rate of RA was 3.75/100,000 doses of quadrivalent HPV. There was no cluster of these adverse events. Considering the prevalence of RA is 0.46% [21] and the lack of large epidemiology studies of ON [22] and HSP in Chinese women, the rate of adverse events was not high enough for a safety signal. Furthermore, an updated systematic review concluded that the risk of RA, ON, or HSP is not increased with any type of HPV vaccine [23]. However, due to the limited sample size and different backgrounds of Chinese women, a larger sample or longer observational period is required to comprehensively review the safety of HPV vaccines in China.

To monitor the long-term safety of HPV vaccines, a continuous, sustainable surveillance system is necessary. There are several ways of setting controls to promptly identify and alert of vaccine safety signals. First, comparing with historical controls, or background rates, is an approach to identify whether the incidence rate of adverse events following vaccination is higher than the background rate, which is not due to chance [24]. However, it may not be feasible to perform this comparison in the Ningbo database due to the short premarketing period of the HPV vaccine. Four study designs were recently recognized and have become widespread for vaccine safety surveillance, including the self-controlled case series (SCCS), self-controlled risk interval (SCRI), cohort, and case-control study designs [25], and SCRI and SCCS have been proven to be an efficient, rational alternative to the cohort in a simulation study [25,26]. In a real-world study, self-controlled analyses, such as those used in SCRI and SCCS, are used because each case acts as its own control within a short period, thereby inherently adjusting for all time-unvarying potential confounders, such as sex, nationality, and genetic predisposition [27,28]. In our study, given that the database can only capture the visit records of safety outcomes and HPV vaccines that occurred in the city, it is likely that we will misclassify these crucial variables in a subsequent study; for example, some local young women can be vaccinated while attending college outside the city. In addition, these records cannot currently be traced back to the Ningbo database unless the women return to Ningbo and utilize the relevant local services.

Therefore, using the SCRI as the primary design to evaluate the risk after vaccination seems most appropriate, minimizing misclassification bias due to incomplete vaccine exposure. A cohort study as the secondary design may improve statistical power [28]. Also, additional regional databases need to be incorporated to increase the robustness of risk estimation, especially for different subpopulations or districts.

Limitations

Some potential barriers need to be overcome in additional studies. First, because the HPV vaccination rate is so low in Ningbo, the observation period should be lengthened, or additional sites should be incorporated. Second, identification algorithms for safety outcomes should be developed to improve performance, with a positive predictive value >70% [29]. Third, migration may cause potential bias in the representative population or misclassification of exposure and outcomes, so we should consider its influence in future research. Finally, missing links between data should be addressed and balanced with privacy concerns [30].

Conclusions

The study presents an available approach to initiate an active surveillance system for adverse events after HPV vaccination, based on a regional health care information platform in China. Utilizing a longer observation period or including additional functional sites is warranted to conduct future hypothesis-generating and hypothesis-confirming studies to address vaccine safety concerns.

Acknowledgments

The study was funded by the Advanced Project of Beijing Natural Science Foundation (BMU2019GJJXK003) and sponsored by Merck Sharp & Dohme Corp (MSD; Beijing, China). The authors had full access to all data in the study and final responsibility for the decision to submit for publication. The authors' views are independent of the official views of the Beijing Natural Science Committee and MSD.

Authors' Contributions

At the time of the research, LZ led the development of the questionnaire and interview guidelines. YY, MR, and ZS were responsible for reviewing these documents. ZL and XG were responsible for data collection, and FT, DY, and LN assisted with the field investigation. This manuscript was written by LZ and reviewed by ZS. The two corresponding authors (ZS and XG) contributed equally to this work.

Conflicts of Interest

None declared.

References

1. Larson HJ, Cooper LZ, Eskola J, Katz SL, Ratzan S. Addressing the vaccine confidence gap. The Lancet 2011 Aug 06;378(9790):526-535. [doi: 10.1016/S0140-6736(11)60678-8] [Medline: 21664679]
2. Larson HJ, Smith DMD, Paterson P, Cumming M, Eckersberger E, Freifeld CC, et al. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. The Lancet Infectious Diseases 2013 Jul;13(7):606-613. [doi: 10.1016/S1473-3099(13)70108-7] [Medline: 23676442]
3. Coombes R. Europe steps up action against vaccine hesitancy as measles outbreaks continue. BMJ 2017 Oct 16;359:1-2. [doi: 10.1136/bmj.j4803] [Medline: 29038341]
4. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]
5. Bonhoeffer J, Kohl K, Chen R, Duclos P, Heijbel H, Heininger U, The Brighton Collaboration. The Brighton Collaboration: addressing the need for standardized case definitions of adverse events following immunization (AEFI). Vaccine 2002 Dec 13;21(3-4):298-302. [doi: 10.1016/s0264-410x(02)00449-8] [Medline: 12450705]
6. Edwards KM, Hackell JM, COMMITTEE ON INFECTIOUS DISEASES, THE COMMITTEE ON PRACTICE AND AMBULATORY MEDICINE. Countering Vaccine Hesitancy. Pediatrics 2016 Sep;138(3):e1-e14 [FREE Full text] [doi: 10.1542/peds.2016-2146] [Medline: 27573088]
7. Crawford NW, Clothier H, Hodgson K, Selvaraj G, Easton ML, Buttery JP. Active surveillance for adverse events following immunization. Expert Rev Vaccines 2014 Feb;13(2):265-276. [doi: 10.1586/14760584.2014.866895] [Medline: 24350637]
8. Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T, et al. The Vaccine Safety Datalink: a model for monitoring immunization safety. Pediatrics 2011 May;127(S1):S45-S53. [doi: 10.1542/peds.2010-1722H] [Medline: 21502240]
9. McNeil MM, Gee J, Weintraub ES, Belongia EA, Lee GM, Glanz JM, et al. The Vaccine Safety Datalink: successes and challenges monitoring vaccine safety. Vaccine 2014 Sep 22;32(42):5390-5398 [FREE Full text] [doi: 10.1016/j.vaccine.2014.07.073] [Medline: 25108215]
10. Le Saux N, Barrowman NJ, Moore DL, Whiting S, Scheifele D, Halperin S, Members of Canadian Paediatric Society/ Health Canada Immunization Monitoring Program-Active (IMPACT). Decrease in Hospital Admissions for Febrile Seizures and Reports of Hypotonic-Hyporesponsive Episodes Presenting to Hospital Emergency Departments Since Switching to Acellular Pertussis Vaccine in Canada: A Report From IMPACT. Pediatrics 2003 Nov;112(5):e348-e353. [doi: 10.1542/peds.112.5.e348] [Medline: 14595075]
11. Moore DL, Le Saux N, Scheifele D, Halperin SA, Members of the Canadian Paediatric Society/Health Canada Immunization Monitoring Program Active (IMPACT). Lack of Evidence of Encephalopathy Related to Pertussis Vaccine: Active Surveillance by IMPACT, Canada, 1993–2002. Pediatr Infect Dis J 2004 Jun;23(6):568-571. [doi: 10.1097/01.inf.0000130075.56368.02] [Medline: 15194842]
12. Amarasinghe A, Black S, Bonhoeffer J, Carvalho SMD, Dodoo A, Eskola J, et al. Effective vaccine safety systems in all countries: a challenge for more equitable access to immunization. Vaccine 2013 Apr 18;31(S2):B108-B114. [doi: 10.1016/j.vaccine.2012.10.119] [Medline: 23598471]
13. Zhang LX, Wang HB, Li QZ, Zhao MH, Zhan QM. Big data and medical research in China. BMJ 2018 Feb 05;360:1-3 [FREE Full text] [doi: 10.1136/bmj.j5910] [Medline: 29437562]
14. Yang Y, Zhou XF, Gao SQ, Lin HB, Xie YM, Feng YJ, et al. Evaluation of Electronic Healthcare Databases for Post-Marketing Drug Safety Surveillance and Pharmacoepidemiology in China. Drug Saf 2018 Jan;41(1):125-137. [doi: 10.1007/s40264-017-0589-z] [Medline: 28815480]

15. Li HL, Lin HB, Zhao HY, Xu Y, Cheng YC, Shen P, et al. Statins use and risk of new-onset diabetes in hypertensive patients: a population-based retrospective cohort study in Yinzhou district, Ningbo city, People's Republic of China. Ther Clin Risk Manag 2018 May 3;14:823-832 [FREE Full text] [doi: 10.2147/TCRM.S158850] [Medline: 29765224]
16. Lin HB, Tang X, Shen P, Zhang DD, Wu JG, Zhang JY, et al. Using big data to improve cardiovascular care and outcomes in China: a protocol for the CHinese Electronic health Records Research in Yinzhou (CHERRY) Study. BMJ Open 2018 Feb 12;8(2):1-11 [FREE Full text] [doi: 10.1136/bmjopen-2017-019698] [Medline: 29440217]
17. Willame C, Baril L, van den Bosch J, Ferreira GLC, Williams R, Rosillon D, et al. Importance of feasibility assessments before implementing non-interventional pharmacoepidemiologic studies of vaccines: lessons learned and recommendations for future studies. Pharmacoepidemiol Drug Saf 2016 Dec;25(12):1397-1406 [FREE Full text] [doi: 10.1002/pds.4081] [Medline: 27601179]
18. Liang XF, Li L, Liu DW, Li KL, Wu WD, Zhu BP, et al. Safety of Influenza A (H1N1) Vaccine in Postmarketing Surveillance in China. N Engl J Med 2011 Feb 17;364(7):638-647. [doi: 10.1056/NEJMoa1008553] [Medline: 21288090]
19. Wu J, Xu FJ, Lu L, Lu M, Miao L, Gao T, et al. Safety and Effectiveness of a 2009 H1N1 Vaccine in Beijing. N Engl J Med 2010 Dec 16;363(25):2416-2423. [doi: 10.1056/NEJMoa1006736] [Medline: 21158658]
20. Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. Pharmacoepidemiol Drug Saf 2012 Jan;21(S1):291-297. [doi: 10.1002/pds.2323] [Medline: 22262619]
21. Li ZG, Yang Y. Rheumatology in China: challenges and development. Rheumatology (Oxford) 2012 Oct;51(10):1733-1734. [doi: 10.1093/rheumatology/kes166] [Medline: 22772322]
22. Yang C, Fang K, Hu YH. [Research status on the epidemiological characteristics of optic neuritis]. Zhonghua Yan Ke Za Zhi 2015 Dec;51(12):946-949. [Medline: 26888277]
23. Phillips A, Patel C, Pillsbury A, Brotherton J, Macartney K. Safety of Human Papillomavirus Vaccines: An Updated Review. Drug Saf 2018 Apr;41(4):329-346. [doi: 10.1007/s40264-017-0625-z] [Medline: 29280070]
24. Black S, Eskola J, Siegrist CA, Halsey N, MacDonald N, Law B, et al. Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. The Lancet 2009 Dec 19;374(9707):2115-2122 PMID: 19880171, PMID: 20362814 [FREE Full text] [doi: 10.1016/S0140-6736(09)61877-8] [Medline: 19880172]
25. Glanz JM, McClure DL, Xu S, Hambidge SJ, Lee M, Kolczak MS, et al. Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. J Clin Epidemiol 2006 Aug;59(8):808-818. [doi: 10.1016/j.jclinepi.2005.11.012] [Medline: 16828674]
26. McClure DL, Glanz JM, Xu S, Hambidge SJ, Mullooly JP, Baggs J. Comparison of epidemiologic methods for active surveillance of vaccine safety. Vaccine 2008 Jun 19;26(26):3341-3345. [doi: 10.1016/j.vaccine.2008.03.074] [Medline: 18462849]
27. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. Stat Med 2006 May 30;25(10):1768-1797. [doi: 10.1002/sim.2302] [Medline: 16220518]
28. Yih WK, Lieu TA, Kulldorff M, Martin D, McMahill-Walraven CN, Platt R, et al. Intussusception Risk after Rotavirus Vaccination in U.S. Infants. N Engl J Med 2014 Feb 06;370(6):503-512 PMID: 24422677, PMID: 24785219. [doi: 10.1056/NEJMoa1303164] [Medline: 24422676]
29. Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. Pharmacoepidemiol Drug Saf 2012 Jan;21(S1):90-99. [doi: 10.1002/pds.2318] [Medline: 22262597]
30. Weber GM, Mandl KD, Kohane IS. Finding the Missing Link for Big Biomedical Data. JAMA 2014 Jun 25;311(24):2479-2780. [doi: 10.1001/jama.2014.4228] [Medline: 24854141]

Abbreviations

ADEM: acute disseminated encephalomyelitis.
AEFI: adverse events following immunization.
CDC: Centers for Disease Control and Prevention.
DS: death surveillance.
EMR: electronic medical record.
GBS: Guillain-Barre syndrome.
HPV: human papillomavirus.
HSP: Henoch-Schonlein purpura.
ICD-10: 10th revision of the International Statistical Classification of Diseases and Related Health Problems.
ID: identity card.
IP: immunization program.
ITP: primary immune thrombocytopenia.
JRA: juvenile rheumatoid arthritis.

MCH: maternal and child health care.
N/A: not available.
NHC: Ningbo Health Commission.
ON: optic neuritis.
RA: rheumatoid arthritis.
RHR: residents’ health records.
SCCS: self-controlled case series.
SCRI: self-controlled risk interval.
SLE: systemic lupus erythematosus.
VSD: Vaccine Safety Datalink.
VTE: venous thromboembolism.

Edited by G Eysenbach; submitted 12.12.19; peer-reviewed by R Poluru, S Mukherjee; comments to author 21.01.20; revised version received 21.03.20; accepted 30.03.20; published 01.06.20.

Please cite as:
Liu Z, Zhang L, Yang Y, Meng R, Fang T, Dong Y, Li N, Xu G, Zhan S
Active Surveillance of Adverse Events Following Human Papillomavirus Vaccination: Feasibility Pilot Study Based on the Regional Health Care Information Platform in the City of Ningbo, China
J Med Internet Res 2020;22(6):e17446
URL: https://www.jmir.org/2020/6/e17446
doi:10.2196/17446
PMID:32234696

Review

Electronic Data Capture Versus Conventional Data Collection Methods in Clinical Pain Studies: Systematic Review and Meta-Analysis

Lindsay A Jibb[1,2], RN, PhD; James S Khan[3,4], MD, MSc; Puneet Seth[5], MD; Chitra Lalloo[1,6], PhD; Lauren Mulrooney[7], BScN; Kathryn Nicholson[8], PhD; Dominik A Nowak[5,6,9], MD; Harneel Kaur[10], BHSc; Alyssandra Chee-A-Tow[1], BSc, MPH; Joel Foster[11], RN, MPA; Jennifer N Stinson[1,2,6], RN-EC, PhD

[1]Child Health Evaluative Sciences, Hospital for Sick Children, Toronto, ON, Canada
[2]Lawrence S Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada
[3]Department of Anesthesia, Mount Sinai Hospital, Toronto, ON, Canada
[4]Department of Anesthesia, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[5]Department of Family Medicine, McMaster University, Hamilton, ON, Canada
[6]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[7]Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada
[8]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada
[9]Department of Family and Community Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[10]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[11]Office of Education, Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Lindsay A Jibb, RN, PhD
Child Health Evaluative Sciences
Hospital for Sick Children
686 Bay Street
Toronto, ON, M5G 0A4
Canada
Phone: 1 416 813 7654 ext 309160
Email: lindsay.jibb@sickkids.ca

Abstract

Background: The most commonly used means to assess pain is by patient self-reported questionnaires. These questionnaires have traditionally been completed using paper-and-pencil, telephone, or in-person methods, which may limit the validity of the collected data. Electronic data capture methods represent a potential way to validly, reliably, and feasibly collect pain-related data from patients in both clinical and research settings.

Objective: The aim of this study was to conduct a systematic review and meta-analysis to compare electronic and conventional pain-related data collection methods with respect to pain score equivalence, data completeness, ease of use, efficiency, and acceptability between methods.

Methods: We searched the Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica Database (EMBASE), and Cochrane Central Register of Controlled Trials (CENTRAL) from database inception until November 2019. We included all peer-reviewed studies that compared electronic (any modality) and conventional (paper-, telephone-, or in-person–based) data capture methods for patient-reported pain data on one of the following outcomes: pain score equivalence, data completeness, ease of use, efficiency, and acceptability. We used random effects models to combine score equivalence data across studies that reported correlations or measures of agreement between electronic and conventional pain assessment methods.

Results: A total of 53 unique studies were included in this systematic review, of which 21 were included in the meta-analysis. Overall, the pain scores reported electronically were congruent with those reported using conventional modalities, with the majority of studies (36/44, 82%) that reported on pain scores demonstrating this relationship. The weighted summary correlation coefficient of pain score equivalence from our meta-analysis was 0.92 (95% CI 0.88-0.95). Studies on data completeness, patient- or provider-reported ease of use, and efficiency generally indicated that electronic data capture methods were equivalent or superior to conventional methods. Most (19/23, 83%) studies that directly surveyed patients reported that the electronic format was the preferred data collection method.

Conclusions: Electronic pain-related data capture methods are comparable with conventional methods in terms of score equivalence, data completeness, ease, efficiency, and acceptability and, if the appropriate psychometric evaluations are in place, are a feasible means to collect pain data in clinical and research settings.

(J Med Internet Res 2020;22(6):e16480) doi:10.2196/16480

KEYWORDS

electronic; data collection; pain; efficiency; systematic review; meta-analysis

Introduction

Background

Pain is an unpleasant sensory and emotional experience that is unique to the individual. It is also a dynamic process and fluctuates in a multidimensional manner across its sensory (eg, intensity, location, duration, etc), evaluative (ie, impact on functioning) and affective (ie, emotional effect) qualities within both the short and long term [1]. Pain is influenced by a variety of biopsychosocial factors, including genetics, mood, emotions, memory, and interpersonal relationships as well as external stimuli such as physical movement [1-3]. The accurate measurement of pain is of utmost importance to clinicians and researchers.

The most commonly used methods of measuring pain within a clinical and research context are self-reported questionnaires. Clinically, pain measurements are generally performed before and after an intervention to assess a patient's response to therapy. These assessments are typically performed using paper-based questionnaires or via face-to-face or telephone-based verbal surveys or interviews. Although widely used, these conventional data collection methods can introduce a number of biases in the collected pain data. In particular, these methods often rely heavily on a patient's recall of their pain symptoms (eg, pain intensity over the preceding week). Unfortunately, the recall of pain is problematic because memories of pain are vulnerable to distortion due to physical and psychological contextual factors and selective coding and retrieval of memories [4,5]. Additional issues with conventional data collection methods include limitations in conducting ecologically valid assessments of pain in the patient's natural environment and social context, logistical challenges for repeated measurements over time, potential burden to patients, clinicians, and researchers, and possibly reduced data quality due to incomplete or back-filled pain diaries [6-8].

The advent of mobile electronic devices has created novel opportunities to collect pain-related data in clinical and research settings. Electronic data collection methods have been used to assess variables related to a variety of conditions, including mood disorders, asthma, tobacco cessation, urinary incontinence, brain injury, diabetes, cancer, and pain [7,9-11]. Specialists in pain medicine have widely advocated for the use of electronic data capture over the past two decades [12,13], and mounting evidence suggests that data collected via electronic methods may be more accurate and contain fewer errors than conventional methods [14,15]. Although randomized controlled trials and observational studies comparing electronic and conventional data collection methods suggest benefits to the use of electronic devices in pain clinical trials, no review providing an overview of these benefits currently exists. Furthermore, with the advent of smartphone-style mobile phones and their nearly ubiquitous use in developed countries [16], electronic data collection methods are becoming more widely available. As such, a review of the literature is needed to understand the potential advantages and disadvantages of collecting pain data using electronic methods.

Objective

We aimed to identify and synthesize data from studies comparing electronic and conventional pain-related data collection methods to describe similarities and differences in pain scores, data completeness, ease of use, efficiency, and acceptability between methods.

Methods

Overview

We developed an internal protocol to guide the conduct of the review and meta-analysis. Reporting is guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [17].

Eligibility Criteria

Criteria for Inclusion in the Systematic Review

To be included in this review, studies must have (1) been published in English, (2) enrolled participants in a clinical study examining an acute or chronic pain-related outcome as reported by participants, (3) used both an electronic data collection method and a conventional form of data collection (ie, paper-based, telephone, or in-person), and (4) collected data on pain score equivalence (including as part of a functional limitation or disease activity measure), data completeness, ease of use, efficiency, or acceptability between collection methods. There were no restrictions on the type of study design (randomized or observational), country of study, or year of publication. Only studies in which the full texts could be retrieved were included in the review.

Criteria for Inclusion in the Meta-Analysis

A subset of studies included in the systematic review was also included in the meta-analysis. These studies reported correlations or measures of agreement (ie, intraclass correlation coefficients [ICCs], Pearson correlations, Spearman rho, and weighted kappa) between patient-reported pain intensity or pain interference (including affect) scores assessed using an electronic and a conventional data capture method. Pain intensity and interference were the focus of the analysis as these constructs are commonly assessed, single-item aspects of both acute and chronic pain and are routinely used to determine treatment effectiveness and guide therapy [18,19]. As recalled pain reports may not be an accurate reflection of the momentary pain experience, we included only studies that compared momentary pain reports. No restrictions were placed on the type of data collection method (eg, mobile phone, computer-based, and tablet), pain assessment instrument (eg, numerical rating scale [NRS]), frequency of data collection, or other pain-related assessments (ie, studies that also assessed constructs such as quality of life or disease activity in addition to pain intensity or interference were included).

Study Selection

We developed a comprehensive search strategy in consultation with a tertiary hospital librarian with expertise in the scientific literature related to digital health. We customized the search strategy to conduct tailored searches of MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials (CENTRAL) from inception until November 19, 2019. Medical Subject Headings (MeSH) keywords in the search included: *pain*, *pain measurement*, *pain threshold*, *pain perception*, *electronics*, *cellular phone*, *computers*, *handheld*, *wireless technology*, *internet*, *computer communication networks*, *mobile applications*, *randomized controlled trial*, *multicenter study*, *observational study*, *humans*, and *prospective studies*. Additional keywords used in the search included: *pain*, *pain reporting*, *personal digital assistant*, *smartphone*, and *prospective study*. An example of the search strategy can be found in Multimedia Appendix 1. We supplemented our search with searches of the author’s own databases of electronic pain assessment studies.

Search results were initially electronically screened for intradatabase and interdatabase duplicates. After the electronic removal of duplicates, titles and abstracts were screened independently by 2 authors using piloted standardized screening forms (all authors involved). Subsequently, the full texts of the included citations were reviewed in duplicate to confirm study inclusion (all authors involved). The kappa statistic was calculated as a metric of screening agreement at the full-text stage. Following the literature-based precedent, we interpreted the kappa as follows: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81-1.00, almost perfect [20]. Disagreements among reviewers about study eligibility were resolved by consensus through discussion by at least three authors.

Data Collection Process

A standard data collection form was created and piloted. Data abstraction occurred independently and in duplicate. Data extracted included study design, sample size, study population, electronic and conventional data collection method, duration of data collection, score equivalence between data capture methods (ie, correlations, score differences, and descriptive reports), data completeness, ease and efficiency of data collection, and patient or participant acceptability. An *a priori* decision was made to not formally assess study quality given the nature of the intervention (ie, data collection method) and the diverse study designs collected in the systematic search.

Data Synthesis

Descriptive statistics (ie, frequencies and percentages) were used to synthesize and present data across all included studies. Meta-analysis was performed to synthesize results related to score equivalence across data capture methods. For the analysis, reported correlation coefficients (or kappa in the case of 2 studies [21,22]) served as effect size indices. In all studies where more than one coefficient for a correlation or measure of agreement between electronic and conventional pain data collection methods was available, we used the average of the coefficients so that a single study did not disproportionately impact the summary effect size. Whenever available, the reported sample size used to produce the score equivalence coefficient was used in the model. In cases where the sample size for the score equivalence analysis was not explicitly mentioned, we used the sample size reported for the entire study. Random-effects models were used to combine data across studies, and the I^2 statistic was used to quantify heterogeneity. The criteria set out by Higgins et al [23] were used to interpret the I^2 statistic; namely, 25%, 50%, and 75% were considered low, moderate, and high heterogeneity, respectively. To further examine the impact of heterogeneity on the results, the standardized residual score (ie, the standardized difference between each study effect size and the weighted mean effect size) for each study was calculated and compared [9]. A conservative cutoff of ±2 was set to examine extreme effect sizes as determined by the standardized residuals. We performed a sensitivity analysis to evaluate any impact of the type of correlation or measure of agreement on the weighted summary correlation. Specifically, following previously used methods, separate meta-analyses were conducted with studies reporting ICC or weighted kappa, which account for covariance and score mean and variability, and studies reporting the more conventional Pearson or Spearman rho coefficients [9]. Possible publication bias was assessed by visual inspection of an asymmetrical funnel plot. To investigate the sources of heterogeneity, we conducted further subgroup analyses. Our subgroup analyses focused on elucidating the impact of (1) the similarity of pain assessment measure between electronic and conventional modalities (ie, same measure or different) and (2) the duration of data collection (ie, once or multiple times). Subgroup analyses by study participant age and pain condition were precluded by the structure of data reported in our included studies. Meta-analysis procedures were conducted using Microsoft Excel (Microsoft Corporation) and Distiller SR Forest Plot Generator (Evidence Partners Inc).

Results

Study Selection

The search strategy identified 4927 studies, of which 183 underwent full-text review and 129 were excluded (Figure 1). The kappa agreement score between appraisers at this stage was 0.69, which indicated substantial agreement. In all, 54 papers reporting on 53 unique studies were included in the qualitative synthesis. Stinson et al [5,24] reported different results from the same study, so were grouped presently for analyses purposes. In all, 21 studies were included in the quantitative synthesis. The number of published studies meeting our inclusion criteria increased steadily over time (Figure 2).

Figure 1. Study selection flowchart.

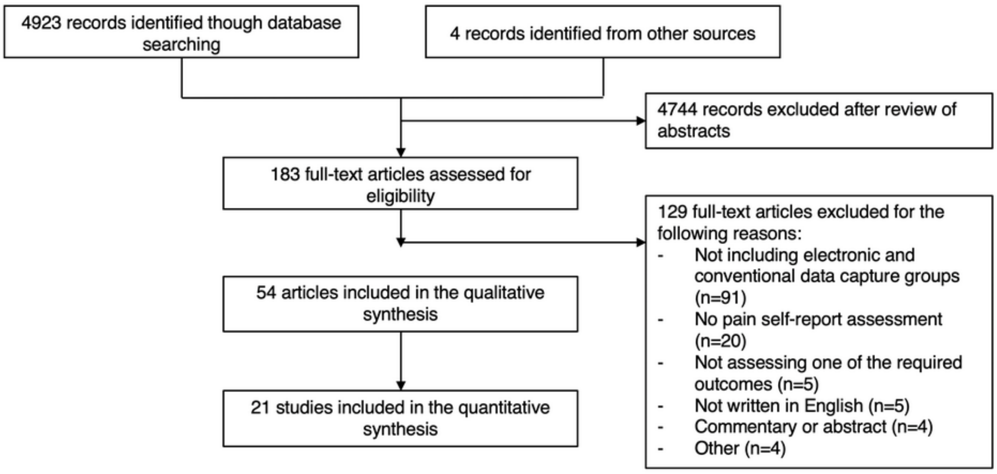
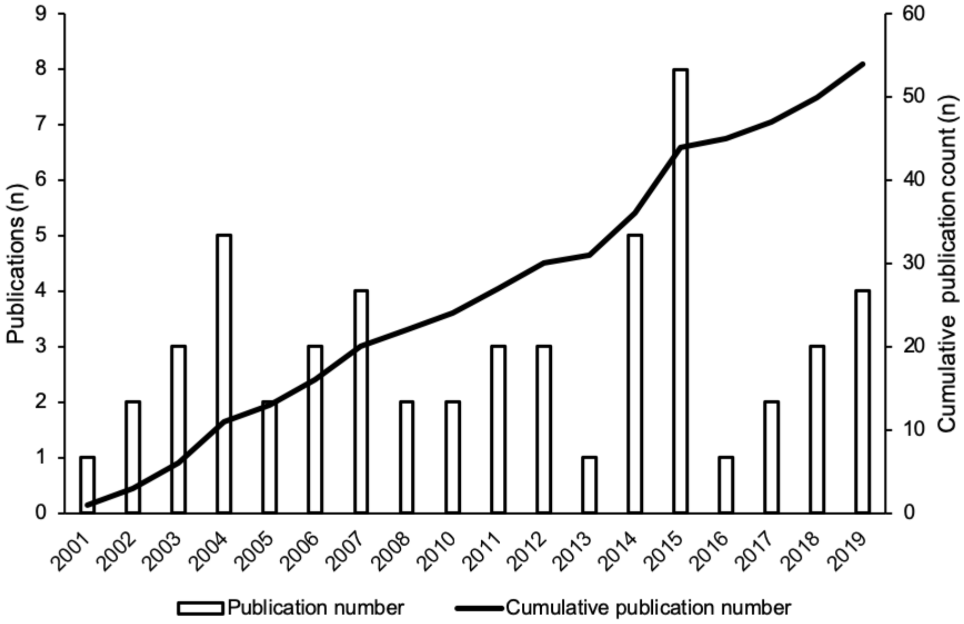


Figure 2. Number of studies meeting inclusion criteria overtime.



Study Characteristics

The study details are presented in Table 1. Data from a total of 7977 pain patients were included in this review. The mean number of participants across studies was 151 (range 15-2400). The average mean or median age of participants was 41.5 years (SD 17.5), and across studies, the average proportion of female participants was 63.1%; mean or median age data were missing from 9 studies and sex data were missing from 7 studies. Participants in the included studies had various painful conditions or diagnoses, including both acute and chronic pain. The most common pain conditions were nonspecific chronic pain (9/54, 17% studies), postoperative pain (8/53, 15% studies), and arthritis (8/53, 15% studies).

Table 1. Study characteristics.

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Allena et al (2012) [25]	Acceptability, data completeness, and ease	Not specified	85	Mean age 39.7 (SD 10.2) years, 68 females and 17 males, medication overuse headache	PDA[a] program collecting data on pain intensity (no indication of measure), pain sensory characteristics, associated symptoms, possible trigger factors and medication use	Paper-based tool (no indication if questions were the same across formats); prospective recording of attack characteristics, more accurate descriptions	Participants completed both formats daily for 7-10 days
Athale et al (2004) [26]	Acceptability, data completeness, ease, and score equivalence	Nonrandomized, crossover	43	Mean age not specified (range 18-75+ years), 36 females and 7 males, rheumatoid arthritis	Computer program collecting data on VAS[b]-rated pain intensity, pain sensory characteristics, and affective and functional impact of pain	Paper-based tool (different from electronic format only in that pain and swelling locations are indicated on separate body maps)	Participants completed each format once
Bandarian-Balooch et al (2017) [27]	Acceptability, data completeness, ease, and score equivalence	Randomized, controlled trial	181	Mean age 26.5 (range 18-55) years, 146 female and 35 males, headache and migraine	Mobile phone or computer program collecting NRS[c]-rated pain intensity, frequency, and duration data as well as triggers and medication use	Paper-based tool with one subgroup identical to electronic format and the other a long-form report representative of conventional paper diaries	Participants completed assigned format once per day for 30 days
Bedson et al (2019) [28]	Data completeness, ease, efficiency, and score equivalence	Nonrandomized, cohort	21	Median age 62 (IQR 50-70) years, 13 females and 8 males, musculoskeletal pain	Tablet program collecting data on NRS-rated pain intensity and pain interference, as well as sleep disturbance, analgesic use, mood, and side effects	Paper-based tool (same assessment as used in the electronic study)	Participants completed electronic assessment 2 times per day for 4 weeks and the paper-based tool once at baseline and once at study completion
Bishop et al (2010) [29]	Acceptability, data completeness, ease, efficiency, and score equivalence	Randomized, crossover	167	Complete age data not reported, (range 18-78), complete sex data not reported, back pain	Computer program collecting data on the occurrence of pain interference (RMDQ[d])	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once in random order on the same day
Blum et al (2014) [30]	Acceptability, ease, and efficiency	Crossover (randomization procedure not stated)	62	Median age 63.5 (range 23-86) years, 31 females and 31 males, cancer	PDA program (E-MOSAIC) collecting data on VAS-rated pain intensity, medication use, and other symptoms	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once with a 1-hour washout between periods

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Byrom et al (2018) [31]	Score equivalence	Randomized, crossover	155	Mean age 48.6 (SD 13.1) years (range 19-69), 83 females and 72 males, chronic pain	Mobile phone or tablet program collecting data on VAS- and NRS-related pain intensity, as well as VRS[e]-rated pain intensity (SF-36[f])	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once with a 30- to 60- min washout between periods
Castarlenas et al (2015) [22]	Acceptability, score equivalence	Crossover (randomization procedure not stated)	191	Mean age 14.6 (range 12-18) years, 117 females and 74 males, pain somewhere in their body in the last 3 months	Mobile phone program collecting data on NRS-rated pain intensity	Verbally administered tool (same assessment as used in the electronic format)	Participants completed each version once
Chiu et al (2019) [32]	Score equivalence	Randomized, crossover	138	Mean age VAS group 55 (SD 14) years, 54 females and 19 males, postoperative pain; mean age NRS group 53 (SD 13) years, 39 females and 26 males, postoperative pain	Mobile phone program collecting data on VAS- and NRS-rated pain intensity	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once with a 5-min washout between periods
Christie et al (2014) [33]	Data completeness and score equivalence	Crossover (randomization procedure not stated)	21	Median age 49.7 (SD 12.2) years, 16 females and 5 males, inflammatory rheumatic disease	Mobile phone program collecting data on NRS-rated pain intensity, fatigue, stiffness and daily activity or function	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format on alternate days for 28 days
Cook et al (2004) [34]	Acceptability, ease, and score equivalence	Randomized, crossover	189	Mean age 47.5 (SD 12.8) years, 119 females and 70 males, chronic pain	Computer program collecting data on VAS- and NRS-rated pain intensity and the affective impact of pain (SF-MPQ[g]). PDI[h] was also used.	Paper-based tool (same assessment as used in the electronic format).	Participants completed both formats once with a 45-min washout between periods
Cunha-Miranda et al (2015) [35]	Score equivalence	Nonrandomized, crossover	134	Mean age 51.3 (SD 12.0) years, 100 females and 34 males, arthritis	Tablet program collecting data on VAS-rated pain intensity and interference, as well as other disease and quality of life metrics dependent on participant diagnosis	Paper-based tool (same assessment as used in the electronic format).	Participants completed each format with a 15-min washout between periods

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Fanciullo et al (2007) [36]	Acceptability and score equivalence	Crossover (randomization procedure not stated)	54	Median age 10.7 (SD 4.0) years, 26 females and 28 males, various causes of pain (eg, broken bones, infections, and cancer)	Computer program collecting data on pain intensity from an investigator-developed computer faces scale	Paper-based tool (Wong-Baker Faces Scale)	Participants completed both formats once
Freynhagen et al (2006) [37]	Ease	Nonrandomized, cohort	717	Mean age 56.0 years (SD not stated), sex ratio not specified, chronic pain	PDA program collecting data on VAS-rated pain intensity, functional disability, and depression	Paper-based tool (same assessment as used in the electronic format)	Participants completed either format once
Gaertner et al (2004) [38]	Acceptability, data completeness, ease, efficiency, and score equivalence	Randomized, crossover	24	Mean age 49.9 (SD 15.1) years, 13 females and 11 males, various painful conditions (eg, cancer, osteoarthritis, chronic neuropathic pain)	PDA program collecting data on NRS-rated pain intensity, analgesic use, other symptoms and therapies	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format daily for 14 days
Garcia-Palacios et al (2013) [39]	Acceptability, data completeness, ease, and score equivalence	Randomized, crossover	47	Mean age 48.1 (SD 8.0) years, 47 females, fibromyalgia	Mobile phone program collecting data on NRS-rated pain intensity, fatigue, and faces scale-rated mood. BPI[i] and fatigue scale were also used.	Paper-based tool (same assessment as used in the electronic format)	Participants completed the electronic assessment 3 times per day for 1 week and the paper-based tool once per week
Heiberg et al (2007) [40]	Acceptability, data completeness, efficiency, and score equivalence	Crossover (randomization procedure not stated)	38	Mean age 58.4 (SD 12.9) years, 25 females and 12 males, rheumatoid arthritis	PDA program collecting data on VAS-rated pain intensity, fatigue, and global disease activity, as well as NRS-rated pain intensity (RADAI[j]) daily, and VRS-rated pain intensity and interference (SF-36) and additional questions on daily functioning collected weekly	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format for 42 days or 6 weeks (21 days/3 weeks for each format)

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Hofstedt et al (2019) [41]	Acceptability and score equivalence	Nonrandomized, cohort	70	Mean age 51.7 (SD 13.2) years, 53 females and 17 males, arthritis	Computer, tablet, or mobile phone program collecting data on VAS-rated pain intensity, global health, and fatigue, as well as disease activity and functional index for a subset of patients	Paper-based tool (same assessment as used in the electronic format)	Participants completed the electronic format at least once during the week before a clinic appointment and the conventional format once at the appointment
Jaatun et al (2014) [42]	Acceptability, ease, score equivalence	Randomized, crossover	92	Age range 20-90 years, 33 females and 59 males, cancer	Tablet program collecting data on pain location from an investigator-developed pain map	Paper-based tool collecting pain location data from the BPI	Participants completed both formats once a 20-30-min washout between periods
Jamison et al (2001) [15]	Data completeness and score equivalence	Nonrandomized, cohort	36	Mean age 42.6 (SD 7.0) years, 20 females and 16 males, chronic low back pain	PDA program collecting data on VAS-rated pain intensity each hour for 16 waking hours as well as number of sleep hours	Paper-based tool collecting data on NRS-rated pain intensity for each waking hour and telephone-based NRS-pain intensity over the preceding week	Participants completed formats for 1 year.
Jamison et al (2002) [43]	Score equivalence	Randomized, crossover	24	Mean age 34.4 (range 19-57) years, 19 females and 5 males, healthy volunteers holding weights heavy enough to induce pain	PDA program collecting data on VAS-rated pain intensity	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format 21 times on 1 day
Jamison et al (2006) [44]	Score equivalence	Nonrandomized, cohort	21	Mean age 42.0 (SD 4.9) years, 9 females and 12 males, low back pain	PDA program collecting data on VAS-rated pain intensity, as well as the affective and functional impact of pain, medications, and side effects	Telephone interviews collecting data on recalled NRS-rated pain over the previous week and telephone-based NRS-pain intensity over the preceding week	Participants completed the electronic format at least daily for 1 year.
Jonassaint et al (2015) [45]	Score equivalence	Nonrandomized, cohort	15	Median age 29 (range 16-54) years, 6 females and 9 males, sickle cell disease	Mobile phone program collecting VAS-rated pain intensity, location and perceived severity, and treatment strategies.	Paper-based tool collecting data on VAS-rated pain (same assessment as used in the electronic format)	Participants first completed paper-based tool, then electronic version daily for 28 days.

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Junker et al (2008) [46]	Data completeness and score equivalence	Randomized, crossover	198	Mean age 56.5 (SD 13.9) years, 114 females and 84 males, chronic pain	PDA program collecting data on VAS-rated pain intensity recalled pain over previous 4 weeks, recalled worst pain in previous 4 weeks and a summative pain score	Paper-based tool (different from electronic format in that pain intensity rated on NRS)	Participants completed each format once
Khan et al (2019) [47]	Acceptability and data completeness	Randomized, cohort	78	Mean age 52.7 (SD 11.1) years, 78 females, postoperative pain	Computer, mobile phone, or tablet program collecting data on data on NRS-related pain intensity, as well as pain catastrophizing, preoperative anxiety, and somatic preoccupation presurgery and medication use and adverse events postsurgery	Paper- or in-person verbal tool (same assessment as used in the electronic format)	Participants completed each format twice daily on postoperative days 1, 2, 3, and 9 and at a 3-month follow-up visit
Kim et al (2016) [48]	Acceptability and efficiency	Nonrandomized, cohort	96	Mean age not specified, 59 females and 37 males, spinal disorders	Tablet program collecting data on VAS-rated pain intensity, disability, as well as questions related to the nature of pain and alleviating and aggravating pain factors	Paper-based tool (same assessment as used in electronic format)	Each format used for a variable and unspecified number of times
Koho et al (2014) [49]	Acceptability, ease, and score equivalence	Randomized, crossover	94	Mean age 47.0 (SD 8.0) years, 55 females and 39 males, chronic musculoskeletal pain	Computer program collecting data on the affective impact of pain	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format twice on two consecutive days

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Kvien et al 2005 [50]	Acceptability, efficiency, and score equivalence	Nonrandomized, crossover	30	Mean age 61.6 (range 49.8-70.0) years, 19 females and 11 males, rheumatoid arthritis	PDA program collecting data on VAS-rated pain intensity, fatigue, and patient global evaluation of their disease, NRS-rated pain intensity (RADAI), VRS-rated pain intensity and interference (SF-36), and additional questions on daily functioning	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format on 2 occasions 5 to 7 days apart
MacKenzie et al (2011) [51]	Acceptability, ease, efficiency, and score equivalence	Randomized, crossover	63	Mean age 53.0 (range 28.0-82.0) years, 29 females and 34 males, psoriatic arthritis	Computer program collecting data on VAS-rated pain intensity (HAQ[k]), VRS-rated pain intensity and interference (SF-36) and additional questions on health and arthritis-related symptoms and function	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once 1 hour apart
Marceau et al (2007) [52]	Acceptability, data completeness, ease and score equivalence	Randomized, crossover	36	Mean age 48.0 (SD 8.0) years, 25 females and 11 males, chronic pain	PDA program collecting data on VAS-rated pain intensity and interference, as well as on the affective impact of pain, medication use, and pain location	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once per day for 2 weeks with a 1-week washout between periods
Marceau et al (2010) [53]	Acceptability and ease	Randomized, controlled trial	134	Mean age 49.5 (SD 11.3) years, 67 females and 67 males, chronic pain	PDA program collecting data on VAS-rated pain intensity and interference, as well as on the affective impact of pain, medication use, and pain location	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format monthly for 10 months
Matthews et al (2018) [54]	Score equivalence	Randomized, crossover	32	Mean age 24.5 (SD 5.6) years, 25 females and 7 males, nontraumatic knee pain	Tablet-based method of collecting data on pain area, location, and distribution through drawing	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once with a 1-2-min washout between periods

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Neudecker et al (2006) [55]	Score equivalence	Randomized, crossover	53	Mean age 51.0 (range 18.0-78.0) years, 33 females and 20 males, postoperative pain	PDA program collecting data on VAS-rated pain intensity	Manually manipulated slide device-based tool (same assessment as used in the electronic format)	Participants completed each format while participants were at rest and while coughing (number of assessments not specified)
Palermo et al (2004) [56]	Acceptability, data completeness, ease, and score equivalence	Randomized, controlled trial	60	Mean age electronic version 12.3 (SD 2.4) years, mean age paper version 12.3 (SD 3.0) years, 42 females and 18 males, headache or juvenile idiopathic arthritis	PDA program collecting data on faces scale-rated pain intensity, pain sensory characteristics, affective and functional impact of pain	Paper-based tool (same assessment as used in the electronic format)	Participants completed the assigned format for 7 consecutive days
Pawar et al (2017) [57]	Acceptability, ease, efficiency, and score equivalence	Randomized, crossover	52	Mean age 46.6 (SD 14.5) years, 31 females and 21 males, low back pain	Mobile phone program collecting data on the occurrence of pain interference (RMDQ)	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format with a 1-hour interval between assessments
Ritter et al (2004) [58]	Data completeness and score equivalence	Randomized, controlled trial	397	Mean age electronic version 45.9 (SD 14.3) years, mean age paper version 44.6 (SD 13.5) years, 287 females and 110 males, diabetes, asthma, heart disease, lung disease, hypertension	Computer program collecting data on 16 health-related variables including NRS-rated pain intensity	Paper-based tool (same assessment as used in the electronic format)	Participants completed assigned format once
Rolfson et al (2011) [59]	Data completeness and score equivalence	Randomized, controlled trial	2400	Group mean age and sex ratio not specified, total hip replacement surgical pain	Computer program collecting data on VAS-rated pain intensity and health-related quality of life	Paper-based tool (same assessment as used in the electronic format)	Participants completed assigned format once
Saleh et al (2002) [60]	Acceptability and score equivalence	Nonrandomized, cohort	87	Mean age 63.5 (SD 11.6) years, 3 females and 84 males, hip or knee pain	PDA program collecting data on VRS-rated pain intensity and interference (SF-36) and NRS-rated pain interference (WOMAC[l])	Paper-based tool (same assessment as used in the electronic format)	Participants completed assigned format once
Sanchez-Rodrıguez et al (2015) [61]	Acceptability and score equivalence	Nonrandomized, crossover	180	Mean age 14.9 (SD 1.64; age range: 12–19) years, 104 females and 76 males, pain in the last 3 months	Mobile phone program, collecting NRS-, faces pain scale-, VAS-and CAS[m]-pain intensity data	Paper-based tool (same assessment as used in the electronic format)	Participants completed each assigned format once with a 30-min interval between assessments

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Serif et al 2005 [62]	Ease and efficiency	Nonrandomized, cohort	50	Age range 27-65 years, sex not specified, back pain	PDA program collecting data on VAS-pain intensity, pain location, and other symptoms	Paper-based tool (same assessment as used in the electronic format)	Participants completed assessments every 2 hours (between 10 am and 4 pm) for 5 days
Stinson et al (2008 and 2014) [5,24]	Acceptability, data completeness, ease, efficiency, and score equivalence	Nonrandomized, cohort	76 in nonjoint injection group and 36 in joint injection group	Mean age nonjoint injection group 13.4 (SD 2.5) years, 59 females and 17 males, arthritis; mean age joint injection group 12.6 (SD 2.4) years, 24 females and 12 males, arthritis	PDA program collecting data on VAS-rated pain intensity, interference and unpleasantness	Paper based tool (different from the electronic tool in that recall period was 1 week) and quality of life and pain coping also assessed	Participants completed the electronic format 3 times daily for 14 days (21 days for joint injection group) and the conventional format on days 7 and 14 (and 21 for joint injection group)
Stinson et al (2012) [63]	Acceptability, data completeness, ease, efficiency, and score equivalence	Randomized, crossover	24 children aged 4-7 years (with parents) and 77 youth aged 8-18 years	Mean age younger children 5.9 (SD 0.9) years, mean age older children 13.5 (SD 3.1) years, 61 females and 36 males, various rheumatic diseases	(1) Mobile phone program collecting data on faces scale or NRS-rated pain intensity, pain sensory characteristics and affective and functional impact of pain and (2) computer program (same assessment as used in the mobile phone format)	Paper-based tool (same assessment as used in the electronic formats)	Participants completed each format once
Stinson et al (2015) [7]	Acceptability, data completeness, ease, efficiency, and score equivalence	Nonrandomized, cohort	92 in nonsurgical group and 14 in surgical group	Mean age nonsurgical group 13.1 (SD 2.9) years, 45 females and 47 males, cancer; mean age surgical group 14.8 (SD 2.8) years, 7 females and 7 males, cancer surgery	Mobile phone program collecting data on VAS-rated pain intensity, interference and unpleasantness, as well as pain duration and location, pain management strategies used	Paper-based tool (different from the electronic tool in that recall period was 1 week) and quality of life and pain coping also assessed	Participants completed the electronic format twice daily for 14 days (21 days for surgical group) and the conventional format on days 7 and 14 (and 21 for surgical group)

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Stomberg et al (2012) [64]	Acceptability, data completeness, ease, efficiency, and score equivalence	Randomized, controlled trial	40	Age range 18-66 years, sex ratio not specified, posthysterectomy and postcholecystectomy pain	Mobile phone program collecting data on NRS-rated pain intensity	Paper-based tool (same assessment as used in the electronic format)	Participants in the electronic group completed pain assessments every 4 hours during the day for 6 days, plus ad hoc reports, participants in the conventional group completed pain assessments every 4 hours during the day for 4 days
Stone et al (2003) [65]	Data completeness and score equivalence	Randomized, controlled trial	91	Mean age across groups 49.0-53.5 (SD 10.4-10.7) years, 77 females and 14 males, chronic pain	PDA program collecting data on VAS-rated pain intensity, pain sensory characteristics, and affective and functional impact of pain	Paper-based tool (same assessment as used in the electronic format)	Participants in the electronic group completed pain assessments either 3, 6, or 12 times per day for 2 weeks, participants in the conventional group completed pain assessments once per week for 2 weeks.
Sun et al (2015) [66]	Acceptability and score equivalence	Randomized, crossover	128	Median age faces pain scale group 7.5 (range 4-12 years), median age CAS group 13 (range 5-18 years), 52 females and 76 males, postoperative pain	Mobile phone program collecting data on faces pain scale- (children <5 years) and CAS- (children 5-12 years) rated pain intensity	Paper-based tool (same assessment as used in the electronic format)	Participants completed each tool within 10 min of waking from surgery and 30 min later with a 5-min washout interval in between
Suso-Ribera et al (2018) [67]	Data completeness, ease, and score equivalence	Nonrandomized, cohort	38	Mean age 42.7 (SD 9.9) years, 20 females and 18 males, chronic pain	Mobile phone-based program collecting data on NRS-rated pain intensity and interference, as well as pain catastrophizing, pain acceptance, and fear and avoidance, mood and coping	Paper- and telephone-based tool collecting data on NRS-rated pain intensity and interference, as well as pain catastrophizing, pain acceptance, and fear/avoidance, mood and coping (tools used may have differed from electronic format)	Participants completed the electronic format twice daily for 30 days and the conventional format at baseline and after each study week

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Symonds et al (2015) [68]	Score equivalence	Nonrandomized, crossover	356	Mean age across groups 58.4 (SD 8.4) years, 279 females and 77 males, osteoarthritis of the index knee	PDA program collecting data on VRS-rated pain intensity and interference (SF-36) and NRS-rated pain interference (WOMAC)	Paper-based tool collected data from the WOMAC	Participants complete each format once (washout period not specified)
Theiler et al (2007) [69]	Acceptability	Nonrandomized, cohort	60	Mean age 52.1 (range 23.0-79.0) years, 36 females and 24 males, chronic pain	Computer program collecting data on NRS-rated pain intensity, medication use, and other symptoms	Telephone-based tool (same assessment as used in the electronic format)	Participants completed either format every day for 1 week followed by 3-4 days per week for 3 additional weeks
VanDenKerkhof et al (2003) [70]	Data completeness, efficiency, and score equivalence	Nonrandomized, cohort	84	Age and sex ratio not specified, postorthopedic surgical pain	PDA-based program collecting data on NRS-rated pain intensity and physician orders	Paper-based tool (same assessment as used in the electronic format)	Physician completed each format for half of the study period, assessments were completed once per participant
VanDenKerkhof et al (2004) [71]	Data completeness and efficiency	Randomized, controlled trial	74	Mean age electronic group 64.0 (SD 10.0) years, mean age conventional group 58.0 (SD 16.0) years, sex ratio not specified, postorthopedic surgical pain	PDA program collecting data on NRS-rated pain intensity and physician orders	Paper-based tool (same assessment as used in the electronic format)	Participants completed assigned format once
Wæhrens et al (2015) [72]	Acceptability, ease, and score equivalence	Randomized, crossover	20	Mean age 47.8 (SD 11.0) years, 20 females, chronic widespread pain	Computer program collecting data on NRS-rated pain intensity, interference, affect as part of the FIQ[n], as well as measures of depression, quality of life, coping and anxiety	Paper based tool (same assessment as used in the electronic format)	Participants completed each format once with a 5-min wash-out interval

Authors (publication year)	Criteria for electronic and conventional pain assessments	Study design	Sample size	Population (age, sex, pain condition)	Electronic data collection modality and pain data collected	Conventional data collection method and pain data collected	Duration of data collection
Wood et al (2011) [21]	Acceptability and score equivalence	Randomized, crossover	202	Mean age 8.3 (SD 2.6) years, 85 females and 117 males, postoperative or disease-related pain	PDA program collecting data on faces scale-rated pain intensity	Paper-based tool (same assessment as used in the electronic format)	Participants completed each format once with a 30-min washout between periods

[a] PDA: personal digital assistant.

[b] VAS: Visual Analog Scale.

[c] NRS: Numerical Rating Scale.

[d] RMDQ: Roland Morris Disability Questionnaire.

[e] VRS: Verbal Rating Scale.

[f] SF-36: Short Form 36 Health Survey.

[g] SF-MPQ: Short Form McGill Pain Questionnaire.

[h] PDI: Pain Disability Index.

[i] BPI: Brief Pain Inventory.

[j] RADAI: Rheumatoid Arthritis Disease Activity Index.

[k] HAQ: Health Assessment Questionnaire.

[l] WOMAC: Western Ontario and McMaster University Osteoarthritis Index.

[m] CAS: Color Analogue Scale.

[n] FIQ: Fibromyalgia Impact Questionnaire.

Regarding electronic data capture modalities, the devices used for data collection included the following: personal digital assistants (PDA; 22/53, 41%), computer (either Web-based or offline; 10/53, 18%), smartphone (9/53, 17%), tablet (5/53, 9%), mobile phones, tablets, and//or computers (6/53, 11%), and conventional mobile phone (1/53, 22%). Studies conducted more recently tended to use non-PDA–based mobile modalities, whereas older studies utilized PDA and computer-based modalities of assessment (average year of publication for studies employing non-PDA mobile devices was 2016 versus 2007 for studies on PDA and computer-based modalities). Conventional pain assessment modalities were paper-based (46/53, 86.7%), telephone-interviews (2/53; 43%), paper- and verbal-based (3/53, 65%), face-to-face interviews (1/53, 22%), and a manually manipulated slide device (1/53, 22%).

In total, 35% (19/53) studies used a randomized, crossover design, 14 (26%) studies used a nonrandomized cohort design, 9 (17%) studies were randomized controlled trials, 5 (9%)studies used a nonrandomized crossover design, 5 (9%) studies used a crossover design with unclear randomization (no mention of whether a randomization procedure was employed), and 1 (22%) study did not specify the study design. The duration of data collection varied across studies, ranging from a single assessment being conducted to repeated assessments over the course of a year.

Data Related to Pain Assessment Measures

Pain intensity was the most commonly assessed pain outcome, measured in 90% (48/53) of studies. Methods to measure pain intensity using electronic methods were visual analog scales (VAS; 26/53, 49%), NRS (22/53, 41%), faces scales (5/53, 9%), verbal rating scales (5/53, 9%), and color analogue scales (2/53, 44%). The method of pain intensity measurement was not specified in 1 study (21.9%). In total, 75% (40/53) of studies employed the same measurement tools across the electronic and conventional modalities.

Pain assessment tools using electronic data capture most often were multidimensional in nature (35/53, 66%). Electronic data collection methods were used to capture multidimensional aspects of pain using the following validated questionnaires: Brief Pain Inventory, Fibromyalgia Impact Questionnaire, Health Assessment Questionnaire, Pain Disability Index, Rheumatoid Arthritis Disease Activity Index, Roland-Morris Disability Questionnaire, Short Form 20, Short Form 36, Short Form McGill Pain Questionnaire, and Western Ontario and McMaster Universities Arthritis Index.

Comparisons Across Data Collection Modalities

Qualitative Synthesis of Score Equivalence

In total, 83% (44/53) of studies reported pain score equivalence between electronic and conventional data capture methods (Table 2). Statistical methods used to compare scores differed between studies: 47% (21/44) of these studies used correlational analyses (ie, ICC, Pearson coefficient, Spearman coefficient, or weighted kappa) to examine the agreement between pain scores; 29% (13/44) studies statistically examined the differences between mean or median score, SDs, or ranges between methods; 76% (3/44) studies used descriptive methods to examine agreement; and 15% (7/44) studies used a combination of these statistical methods.

Table 2. Summary of study results related to score equivalence.

Outcome and study (year)	Equivalence examination method and results				
	Score correlation		Score differences		Descriptive
	Method	Results	Method	Results	
Studies reporting pain score equivalence					
Athale et al (2004) [26]	ICC[a]	Pain intensity ICC=0.941; pain interference ICC=0.959	—[b]	—	—
Bandarian-Balooch et al (2017) [27]	—	—	ANOVA[c]	Mean pain intensity, frequency, duration, medication usage, disability *P*>.05 of all	—
Bishop et al (2010) [29]	ICC	Pain interference ICC=0.965	—	—	Mean low-back pain interference score difference between method 0.03 (SD 1.43; 95% CI −0.19 to 0.25). Authors predefined acceptable 95% CI was ± 0.5.
Byrom et al (2018) [31]	ICC	Pain intensity *r*=0.87-0.98, 95% CI 0.83-0.99)	—	—	—
Castarlanas et al (2015) [22]	Weighted kappa	Pain intensity κ=0.813	—	—	—
Chiu et al (2019) [32]	Pearson correlation	Pain intensity *r*=0.93-0.96 (*P*<.001)	—	—	Using Bland-Altman method, an agreement between the data capture techniques shown at 95% CI.
Christie et al (2014) [33]	—	—	Paired sample *t* tests or Wilcoxon Signed Rank Test	Mean, SD, and range of pain intensity *P*>.46 for all	—
Cook et al (2004) [34]	Spearman rho	Pain intensity and interference rho=0.67-084	—	—	—
Cunha Miranda et al (2015) [35]	ICC	Pain intensity and interference ICC=>0.781-0.944	—	—	—
Fanciullo et al (2007) [36]	Spearman rho	Pain intensity rho=−0.72 (*P*<.001)	—	—	—
Gaertner et al (2004) [38]	—	—	*t* test	Mean pain intensity not significantly different (*P* value not reported)	—
Garcia-Palacios et al (2013) [39]	Pearson correlation	Pain intensity *r*=0.79 (*P*<.001)	—	—	—
Heiberg et al (2007) [40]	—	—	Wilcoxon's signed rank test	Mean, SD, and range of pain intensity *P*>.06	—
Hofstedt et al (2019) [41]	ICC	Pain intensity ICC=0.952	Paired *t* test	Mean pain intensity not significantly different (*P*=.29)	Using Bland-Altman method, an agreement between the data capture techniques shown at 95% CI.
Jaatun et al (2014) [42]	—	—	—	—	In 71% (65/92) of cases participants marked the same number of areas and the same anatomical locations on both body map versions, in 20 cases, the markings were relatively similar, and in 7 cases, the markings were dissimilar.

Outcome and study (year)	Equivalence examination method and results				
	Score correlation		Score differences		Descriptive
	Method	Results	Method	Results	
Jamison et al (2001) [15]	Pearson correlation	Pain intensity r=0.88, P<.001	—	—	—
Jamison et al (2002) [43]	Pearson correlation	Pain intensity r^2>0.999	—	—	—
Jamison et al (2006) [44]	Pearson correlation	Pain intensity r=0.99 (95% CI 0.975-0.996)	—	—	—
Jonassaint et al (2015) [45]	ICC	Pain intensity ICC=0.97 (95% CI 0.88-0.99)	—	—	—
Kvien et al (2005) [50]	Pearson correlation	Pain intensity r=0.79-0.93	—	—	—
MacKenzie et al (2011) [51]	ICC	Pain intensity and interference ICC=0.95-0.97; 95% CI 0.95-0.98)	—	—	—
Marceau et al (2007) [52]	—	—	—	—	Participants reported similar using each data capture methods for pain intensity, pain interference, mood, and helpfulness of medications.
Matthews et al (2018) [54]	Pearson correlation and ICC	Pain location pixelated area r=0.93 (P<.001) and ICC=0.966 (P<.001)	t test	Mean pain location pixelated area not significantly different (P=.93)	Using Bland-Altman method, an agreement between the data capture techniques shown at 95% CI.
Neudecker et al (2006) [55]	Pearson correlation	Pain intensity r=0.902 (P<.001)	—	—	—
Palermo et al (2004) [56]	—	—	t test	Mean pain intensity not significantly different (P value not reported)	—
Pawar et al (2017) [57]	ICC	Pain interference ICC=0.994 (95% CI 0.989-0.996)	—	—	—
Ritter et al (2004) [58]	—	—	t test, Wilcoxon's signed rank test and ANCOVA[d]	Mean pain intensity and pain interference P>.30	—
Saleh et al (2002) [60]	—	—	Test not reported	Mean and SD pain intensity and interference not significantly different (P value not reported)	—
Sanchez-Rodrıguez et al (2015) [61]	—	—	—	—	Using Bland-Altman method, an agreement between the data capture techniques shown for the FPS-R[e], the VAS[f], and the CAS[g] at 95% CI. Agreement for the NRS[h]-11 shown in the 80% CI level.
Stinson et al (2012) [63]	—	—	t test	Mean pain intensity P>.09 for younger and older children	—
Stinson et al (2015) [7]	Pearson correlation	Pain intensity r=0.49-0.63 (P<.001); pain interference r=0.53-0.65 (P<.001)	—	—	—

Outcome and study (year)	Equivalence examination method and results				
	Score correlation		Score differences		Descriptive
	Method	Results	Method	Results	
Stone et al (2003) [65]	—	—	Repeated-measures ANOVA	Mean pain intensity *P*>.16	—
Sun et al (2015) [66]	Pearson correlation	Pain intensity *r*=0.87-0.93	—	—	Using Bland-Altman method, agreement between the data capture techniques shown in the 80% CI level.
Symonds et al (2015) [68]	Pearson correlation and ICC	Pain intensity *r*=0.92 and ICC=0.92; pain interference *r*=0.97 and ICC=0.97	—	—	—
VanDenKerkhof et al (2003) [70]	—	—	Mann-Whitney test	Median pain intensity not significantly different (*P* value not reported)	—
Wood et al (2011) [21]	Weighted kappa and Spearman rho	Pain intensity κ 0.846 (95% CI 0.79-0.896) and rho=0.911 (*P*<.001)	—	—	—
Studies reporting pain score nonequivalence					
Rolfson et al (2011) [59]	—	—	Mann-Whitney *U* test	Mean pain intensity *P*=.02	—
Studies reporting discrepant results					
Bedson et al (2019) [28]	Spearman rho	Pain intensity and interference baseline paper-based and first 3 days of electronic reports rho=0.60 −0.79 (*P*<.006); pain intensity and interference last 3 days of electronic reports and follow-up paper-based rho=0.40 (*P*<.11)-0.92 (*P*<.001)	—	—	—
Junker et al (2008) [46]	—	—	Paired *t* test	Mean average and present pain intensity *P*<.01; mean worst pain *P*=.68 (null hypothesis was nonequivalence)	—
Koho et al (2014) [49]	ICC	Pain-related fear ICC=0.77 (95% CI 0.66-0.85)	Test not reported	Significantly higher mean scores for 2 of 17 scale items using the electronic method (*P* value not reported)	Using Bland-Altman method, an agreement between the data capture techniques shown at 95% CI.
Stinson et al (2008 and 2014) [5,24]	Pearson correlation and ICC	Pain intensity *r*=0.55-0.76 and ICC=0.52-0.75 (*P*<.01); pain interference *r*=0.77-0.84 (*P*<.01)	—	—	—
Stomberg et al (2012) [64]	—	—	Mantel's test	Mean pain intensity significantly higher in electronic data capture group on 2 of 3 assessment days (*P* value not reported)	—

Outcome and study (year)	Equivalence examination method and results				
	Score correlation		Score differences		Descriptive
	Method	Results	Method	Results	
Suso-Ribera et al (2018) [67]	Pearson correlation	Pain intensity and interference r=0.60-0.81	Paired sample t tests	Averaged weekly pain interference reports from app significantly lower than verbally or paper-based recalled interference verbal over the week P<.001	—
Wæhrens et al (2015) [72]	ICC	Pain intensity and pain interference ICC=0.76-0.98 (95% CI 0.50-0.99)	—	—	—

[a] ICC: intraclass correlation coefficient.

[b] N/A: not applicable.

[c] ANOVA: analysis of variance.

[d] ANCOVA: analysis of covariance.

[e] FPS-R: Faces Pain Scale-Revised

[f] VAS: Visual Analog Scale.

[g] CAS: Color Analogue Scale.

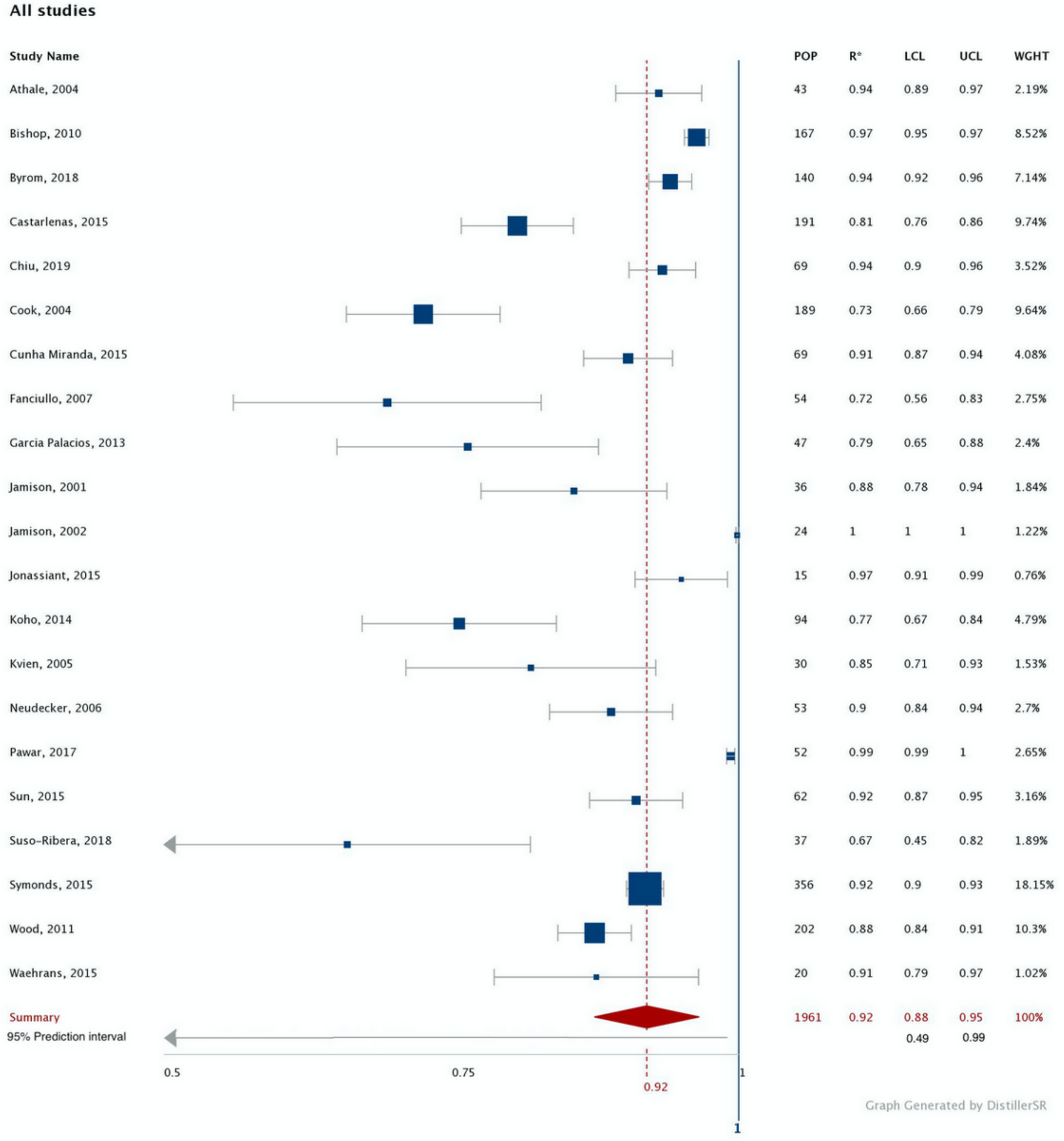
[h] NRS: Numerical Rating Scale.

Across all methods used to compare scores, 82% (36/44) studies demonstrated equivalence between scores reported electronically or using conventional methods. One of these 44 studies (2%) reported nonequivalent scores between data collection methods, and 16% (7/44) studies reported discrepant results. Among studies reporting nonequivalence or discrepancies, purported reasons were recall bias, differences in question layout wherein paper assessments made all items visible to participants simultaneously allowing item scoring in relation to other responses, capacity to change item response using paper methods, and differences in scale presentation (eg, numerical values for NRS not shown using electronic data capture method).

Quantitative Synthesis of Score Equivalence

A forest plot for correlations for score equivalence between data collection modalities is shown in Figure 3. The weighted summary correlation coefficient was 0.92 (95% CI 0.88-0.95, n=1961) and considerable heterogeneity (I^2=95%) was observed across studies. Studies using ICC or weighted kappa produced summary correlations that were similar in magnitude to those using Pearson or Spearman rho correlations (ie, 0.91, 95% CI 0.90-0.92, n=1360, I^2=95%; and 0.85, 95% CI 0.82-0.87, n=1159, I^2=95%, respectively). One study met our predefined criterion for extreme effect size [43]. Removing this study from the analysis did not substantially decrease the heterogeneity (I^2=94%), and the summary correlation was essentially unchanged at 0.90 (95% CI 0.86.0.93, n = 1937). Visual inspection of the funnel plot showed asymmetry, suggesting a possible publication bias (Multimedia Appendix 2).

Figure 3. Summary correlation coefficient for pain intensity and interference data collected via electronic and conventional data capture methods (The I^2 and *P* values for heterogeneity are 95% and <0.00001 respectively; the Z and *P* values for the overall effect are 14.4 and <0.00001 respectively; POP: population; R*: correlation coefficient; LCL: lower confidence interval limit; UCL: upper confidence interval limit; WGHT: weight).



All studies

Study Name	POP	R*	LCL	UCL	WGHT
Athale, 2004	43	0.94	0.89	0.97	2.19%
Bishop, 2010	167	0.97	0.95	0.97	8.52%
Byrom, 2018	140	0.94	0.92	0.96	7.14%
Castarlenas, 2015	191	0.81	0.76	0.86	9.74%
Chiu, 2019	69	0.94	0.9	0.96	3.52%
Cook, 2004	189	0.73	0.66	0.79	9.64%
Cunha Miranda, 2015	69	0.91	0.87	0.94	4.08%
Fanciullo, 2007	54	0.72	0.56	0.83	2.75%
Garcia Palacios, 2013	47	0.79	0.65	0.88	2.4%
Jamison, 2001	36	0.88	0.78	0.94	1.84%
Jamison, 2002	24	1	1	1	1.22%
Jonassiant, 2015	15	0.97	0.91	0.99	0.76%
Koho, 2014	94	0.77	0.67	0.84	4.79%
Kvien, 2005	30	0.85	0.71	0.93	1.53%
Neudecker, 2006	53	0.9	0.84	0.94	2.7%
Pawar, 2017	52	0.99	0.99	1	2.65%
Sun, 2015	62	0.92	0.87	0.95	3.16%
Suso-Ribera, 2018	37	0.67	0.45	0.82	1.89%
Symonds, 2015	356	0.92	0.9	0.93	18.15%
Wood, 2011	202	0.88	0.84	0.91	10.3%
Waehrans, 2015	20	0.91	0.79	0.97	1.02%
Summary	1961	0.92	0.88	0.95	100%
95% Prediction interval			0.49	0.99	

Most studies used the same measure (n=16) versus a different measure (n=5) to assess pain via electronic and conventional modalities, and heterogeneity was high in both subgroups. The summary correlation was 0.93 in studies using the same measure (95% CI 0.89-0.96, n=1475, I^2=96%, 95% prediction interval=0.45-0.99) and 0.86 in studies using different measures (95% CI 0.74-0.93, n=526, I^2=90%, 95% prediction interval −0.01-0.99). In the case of data collection duration, 14 studies collected pain data from participants once and 7 collected data on multiple occasions. The summary correlation was 0.92 in studies that collected pain data once (95% CI 0.88-0.95, n=1678, I^2=95%, 95% prediction interval 0.57-0.99) and 0.92 in studies that collected pain data from participants more than once (95% CI 0.75.0.98, n=283, I^2=96%, 95% prediction interval −0.61-0.99). Heterogeneity remained high despite stratification by the duration of data collection.

Data Completeness

Overall, 45% (24/53) studies reported the completeness of data collected via electronic or conventional methods (Table 3). All of these studies compared an electronic data capture modality to paper-based assessments with 8% (2/24) paper-based assessments being mailed to participants. The assessment of data completeness differed across studies and was largely defined as either the percentage of study participants not completing pain assessments or the percentage of missing or incomplete pain assessments. In total, 37% (9/24) studies reported superior data completeness in the electronic data capture group, 33% (8/24) studies reported superior data completeness in the conventional data capture group, 8% (2/24) studies reported mixed results, and 20% (5/24) studies did not conduct a direct comparison between data collection modalities, but reported a high data completeness using electronic data capture.

Table 3. Summary of study results related to data completeness.

Authors (year)	Electronic data collection modality	Conventional data collection modality	Definitions
Allena et al (2012) [25]	Complete records: 98%	Not reported	Defined as the percent of participants completing all assessments
Athale et al (2004) [26]	Missing data: 7/63 (11%)	Missing data: 16/63 (25%)	Defined as the percent of participants completing assessments
Bandarian-Balooch et al (2017) [27]	—[a]	Long-paper diaries had significantly higher missing data scores in data completion than the e-diaries and short-paper diaries ($P<.05$). The short-paper diary had significantly more missing data than the mobile phone groups ($P<.05$) but was not significantly different than the computer group.	Defined as the number of missing items irrespective of inaccurate completion
Bedson et al (2019) [28]	Recordings were made on 73.3% of days	Not reported	Defined as percentage of days on which participants recorded data
Bishop et al (2010) [29]	Missing data: 15 responses (0.004% of items)	Missing data: 3 responses (0.0007% of items)	Defined as the total number of missed assessment items across all participants
Christie et al (2014) [33]	Response rate: 97.9%	Not reported	Defined as the percent of possible text message–based pain assessments completed cross all participants
Gaertner et al (2004) [38]	Missing data: 8% of all daily assessments	Missing data: 0% (participants reported retrospectively completing assessments when they forgot to do so at the scheduled time)	Defined as the percent assessments not completed across all participants over 14 days
Garcia-Palacios et al (2013) [39]	Complete records: 18.2 (86.66%)	Complete records: 11.1 (52.95%; $P<.01$)	Defined as mean number of complete assessments across participants out of possible records
Heiberg et al (2007) [40]	Median value for missing daily data entries: 1 for both periods	Median value for missing daily data entries: 0 for both periods	Defined as median number of missing assessments over 21 days
Jamison et al (2001) [15]	Compliance with reporting: 89.9%	Compliance with reporting: 55.9%	Defined as percent of assessments completed each day for 1 year (365 days; electronic assessments) and percent of assessments completed for 7 days each month for 1 year (84 days; conventional assessment)
Junker et al (2008) [46]	Not reported	Noticeably more missing data on the conventional method when compared with the electronic pain assessment	Defined as number of missing items across each assessment
Khan et al (2019) [47]	Mean number of queries: 1.53 (2.70)	Mean (SD) number of queries: 0.90 (0.87)	Defined as concerns about a specific data point raised by the data manager or study coordinator relating to inappropriate or missing data
Marceau et al (2007) [52]	Complete records: 397/461 (86.1%)	Complete records: 583/583 (100%)	Defined as the number of assessments completed across all participants
Palermo et al (2004) [56]	Compliance: 83.3%	Compliance: 46.7% ($P<.001$)	Defined as the percent of assessments completed over the 7 days
Ritter et al (2004) [58]	Response rate: 87.5%	Response rate: 83.1% ($P=.19$)	Defined as percent of participants who completed assessments
Rolfson et al (2011) [59]	Response rate: 49%	Response rate: 92% ($P<.01$)	Defined as percent of participants who completed assessments
Stinson et al (2008 and 2014) [5,24]	Response rate: 78% and 73% for 2- and 3-week study protocols, respectively	Response rate: 93% in week 1 and 92% in week 2 (not reported for 3-week protocol)	Defined as 100% when 3 diary entries were completed for each of the 14 or 21 days of data collection

Authors (year)	Electronic data collection modality	Conventional data collection modality	Definitions
Stinson et al (2012) [63]	Missing data using Mobile phones: 5.26% (younger children), 3.42% (older children); missing data using computer: 0% (younger children), 0.14% (older children)	Missing data: 0% (younger children), 1.16%/77 (older children; *P*=.047)	Defined as the percent of assessment items not answered by participants
Stinson et al (2015) [7]	Response rate: 72.2% and 47.1% for 2- and 3-week study protocols, respectively	Not reported	Defined as 100% when participants completed 2 diary entries per day for 14 days
Stomberg et al (2012) [64]	Response rate on the day of surgery: 35%; response rate on days 2-4 postoperatively: 100%; response rate on days 5-6 postoperatively: 69%	Response rate on the day of surgery: 41%; response rate on days 2-4 postoperatively: 100%; not required to complete questionnaire on days 5-6	Defined as the percent of participants completing assessments
Stone et al (2003) [65]	Response rate 3 prompts per day: 93.5%; response rate 6 prompts per day: 93.9%; response rate 12 per day 95.5%	Response rate: 100.0%	Defined as the percent of participants completing assessments
Suso-Ribera et al (2018) [67]	Response rate: 75.7%	Not reported	Defined as the percent of completed assessments out of all possible assessments
VanDenKerkhof et al (2003) [70]	NRS[b] score documentation rate: 100%	NRS score documentation rate: 90-97%	Defined as the percentage of time an NRS score was documented during a patient encounter
VanDenKerkhof et al (2004) [71]	Complete records pain scores: 64.7%; complete records nausea, pruritis and sedation side effects: 100%; complete records hypotension side effect: 20.6%	Complete records pain scores: 43.6% (*P*=.07); complete records nausea, pruritis and sedation side effects: 12.8-33.3% of paper assessments (*P*=<.001); complete records hypotension side effect: 5.1% (*P*=.07)	Percent of assessments where outcome was recorded

[a] N/A: not applicable.

[b] NRS: Numerical Rating Scale.

Ease of Use

The ease of use of electronic and/or conventional pain data capture methods was reported in 45% (24/53) studies (Table 4). Ease was assessed subjectively using administered quantitative or qualitative surveys or verbal reports in all studies. Overall, electronic data collection modalities were considered easy to use by patients in pain or their care providers. In 91% (22/24) of the studies, the electronic modality was considered easy to use, easy to understand, or easy to review or report pain. In all, 29% (7/24) studies conducted inferential testing comparing ease between pain data capture modalities. Of these studies, 57% (4/7) showed that electronic versions were significantly easier to use, 14% (1/7) study showed that the paper version was significantly easier to use, and 28% (2/7) studies showed no significant differences between groups.

Table 4. Summary of study results related to ease of use.

Study (year)	Electronic data collection modality	Conventional data collection modality	Conclusion
Allena et al (2012) [25]	Easy to understand: mean 8.7/10; easy to use: mean 8.9/10	Easy to understand: mean 8.3/10; easy to use: mean 7.9/10	Electronic format significantly (*P*<.01) easier.
Athale et al (2004) [26]	9/19 (47%) rated computer as easier	5/19 (26%) rated paper as easier	Not reported
Bandarian-Balooch et al (2017) [27]	Ease of use (all electronic methods combined): mean 6.58/10	Ease of use: mean 6.17/10	The long-paper diary was rated as significantly (*P*<.02) less easy to use than the other diaries
Bedson et al (2019) [28]	100% reported easy to read	Not reported	Not reported
Bishop et al (2010) [29]	17 comments on easy completion	16 comments on easy completion	Not reported
Blum et al (2014) [30]	79% reported no difficulty with using electronic method	Not reported	Not reported
Cook et al 2004 [34]	39% of patients stated easier to understand and complete	24% of patients stated easier to understand and complete	Not reported
Freynhagen et al (2006) [37]	No issues with the use of the PDA[a]	Not reported	Not reported
Gaertner et al (2004) [38]	54% found more complicated	42% found more complicated	No significant difference between modalities
Garcia-Palacios et al (2013) [39]	15/40 (37%) rated easier to use	4/40 (10%) rated easier to use	Not reported
Jaatun et al (2014) [42]	Both physicians found electronic pain reports easier to read and evaluate than the paper maps.	Not reported	Not reported
Koho et al (2014) [49]	64/93 (69%) rated easy to complete, 10/93 (11%) rated difficult to complete	63/93 (68%) rated easy to complete, 10/93 (11%) rated difficult to complete	Not reported
MacKenzie et al (2011) [51]	54/63 (85.7%) rated easy to complete	Not reported	Not reported
Marceau et al (2007) [52]	32/36 (89%) rated easy to understand and use; 30/36 (83%) rated easy to record data	27/36 (75%) rated easy to understand and use; 3/36 (8%) rated easy to record data	No significant difference in ease of understanding and use. Significantly (*P*<.001) higher ease of recording data rating for electronic modality.
Marceau et al (2010) [53]	29/43 (67.4%) rated easy to use and understand	32/35 (91.4%) rated easy to use and understand	Significantly (*P*=.01) higher ease of use and understanding for paper modality.
Palermo et al (2004) [56]	15/18 (83%) rated easy or very easy to remember to fill out	8/15 (53%) rated easy or very easy to remember to fill out	No significant difference between modalities
Pawar et al (2017) [57]	70.58% rated as easy to use	Not reported	Not reported
Serif et al (2005) [62]	Some users, especially those with arthritis and/or poorer eyesight encountered difficulties in using the electronic modality, but ease of use was general consensus	Not reported	Not reported
Stinson et al (2008 and 2014) [5,24]	Majority found the electronic format easy to use	Not reported	Not reported
Stinson et al (2012) [63]	19/21 (91%) of parents the computer or paper to be easier to understand than the handheld device	Not reported	Significant difference (*P*=.03) in opinion of ease of use
Stinson et al (2015) [7]	94.6% and 91.7% of participants in the 2- and 3-week studies, respectively, found electronic diary interfered only minimally with activities	Not reported	Not reported
Stomberg et al (2012) [64]	Mean difficulty in using electronic modality: 1.31/10	No difficulties with use described	Not reported

Study (year)	Electronic data collection modality	Conventional data collection modality	Conclusion
Suso-Ribera et al (2018) [67]	100% of participants found the app extremely easy to use	Not reported	Not reported
Wæhrens et al (2015) [72]	Not reported	None found paper easier to use	Not reported

[a] PDA: personal digital assistant.

Efficiency

In total, 30% (16/53) studies reported on the time to complete pain assessments (Table 5). In all, 44% (7/16) of these studies provided some evidence that pain assessments completed via the electronic modality were quick to complete; 19% (3/16) of these studies provided some evidence that conventional methods to assess pain were quicker; and 1 of 16 studies (6%) showed mixed results where differences in between-assessment modality completion times differed by participant group (eg, older children, parents, and younger children). In all, 25% (4/16) studies indicated that there were no differences in time to complete assessments across methods. Overall, in studies that directly measured the time to complete pain assessments [28,50,51,57,62,63,70,71], the difference in mean times to complete assessments was minimal (ie, <5.6 min).

Table 5. Summary of study results related to efficiency.

Study	Electronic data collection modality	Conventional data collection modality	Study author conclusions
Bedson et al (2019) [28]	Mean and max times to complete pain assessment: 2 and 5 min	Not reported	Not reported
Bishop et al (2010) [29]	19 comments on quick to complete	9 comments on quick to complete	Not reported
Blum et al (2014) [30]	70% completed pain assessment in under 5 min	88% completed pain assessment in under 5 min (questionnaire had fewer times than electronic modality)	Not reported
Gaertner et al (2004) [38]	No difference in time to complete pain assessments between groups (always less than 15 min/day)	—[a]	Not reported
Heiberg et al (2007) [40]	Time to complete the pain assessment similar between groups	—	Not reported
Kim et al (2016) [48]	68.7% responded that the time to complete pain assessments *positive* or *very positive*	Not reported	Significant relationship regarding participants evaluation of the time to complete electronic questionnaire $P<.001$
Kvien et al (2005) [50]	Mean (SD) time to complete pain assessment: 30.5 (16.0) min	Mean (SD) time to complete pain assessment: 24.9 (27.0) min	No significant difference between groups ($P=.11$)
MacKenzie et al (2011) [51]	Mean time to complete pain assessment: 25.0 min (range 5 to 80 min)	Mean time to complete pain assessment: 24.2 min (range 5 to 60 min)	Not reported
Pawar et al (2017) [57]	Mean time to complete pain assessment: 1.28 min (range 0.83-2.63 min)	Mean time to complete pain assessment: 3.7 min (range 2.42-5.23 min)	Not reported
Serif et al (2005) [62]	Mean time to complete pain assessment: 47 seconds	Mean time to complete pain assessment: 267 seconds	Not reported
Stinson et al (2008 and 2014) [5,24]	Most adolescents found the app quick to complete	Not reported	Not reported
Stinson et al (2012) [63]	Computer: mean (SD) time to complete pain assessment: 3.40 (1.53) min for older children, 4.00 (1.71) min for parents and 1.64 (1.50) min for younger children; Mobile phone: mean (SD) time to complete pain assessment: 5.90 (2.79) min for older children, 7.00 (4.08) min for parents and 1.82 (1.17) min for younger children	Mean (SD) time to complete pain assessment: 3.08 (1.66) min for older children, 2.28 (1.32) min for parents and 1.91 (1.81) min for younger children	Completion times significantly longer in electronic group for older children and parents ($P=.001$). No significant difference for younger children ($P=.64$) who completed a shorter assessment.
Stinson et al (2015) [7]	93.2% and 91.7% of participants in the 2- and 3-week studies, respectively, found electronic diary quick to complete	Not reported	Not reported
Stomberg et al (2012) [64]	Participants reported electronic modality not time consuming	Not reported	Not reported
VanDenKerkhof et al (2003) [70]	Median (IQR) time to complete pain assessment: 206 (70) seconds	Median (IQR) time to complete pain assessment: 153 (85) seconds	Completion time significantly longer time to complete using electronic modality ($P<.001$)
VanDenKerkhof et al (2004) [71]	Median (IQR) time to complete pain assessment 2.8 min	Median (IQR) time to complete pain assessment 2.7 min	No significant difference between groups ($P=.74$)

[a] N/A: not applicable.

Acceptability

Data related to the comparative acceptability of each pain assessment modality were collected in 60% (32/53) studies [5,7,21,22,24-27,29,30,34,36,38-42,47-53,56,57,60,61,63,64,66,69,72]. Overall, electronic programs to assess pain are highly acceptable to patients. In total, 19 (83%) of the 23 studies [21,22,25,26,30,34,36,38-42,49-51,57,60,72,73] that directly surveyed patients reported that the electronic format was the preferred data collection method, compared with 1 of 23 studies (4%) [69], which reported that the conventional data collection method was preferable. This study indicated that age was related to patient preference, with younger patients (mean age 45 years) tending to prefer the internet and older patients (mean age 54 years), preferring the telephone-based data collection; 9% (2/23) studies reported discrepant results [66]. One of these studies reported that children aged <8 years favored the electronic assessment method, whereas the parents of these children and children aged 8 to 18 years had no preference. The other study reported that the preferred modality differed depending on the type of pain measurement instrument used. One study (4%) found no difference in participant satisfaction between electronic and conventional pain instruments [47]. Nine studies did not ask patients to specifically declare a preference for assessment modality but still reported high patient satisfaction with the electronic method [5,7,27,29,48,52,53,56,64,74].

Discussion

Principal Findings

This is the first systematic review and meta-analysis to compare electronic and conventional data collection methods for pain-related outcomes. The results from our review suggest strong correspondence in pain scores collected across electronic and conventional modalities as well as ease of use and acceptability for electronic data capture methods. Comparisons of data completeness and efficiency showed mixed results in terms of the superiority of electronic modalities over conventional methods. Overall, these results indicate that electronic data capture is a viable means to assess pain and has the potential to overcome many of the known limitations associated with conventional methods.

The capacity to obtain equivalently scored data from patients across electronic and conventional data capture modalities is paramount to the use of more novel collection methods in clinical and research settings. Studies included in this review (ie, in 82% of cases) commonly reported on the correspondence of pain scores between assessments. Regardless of whether the data analyses were qualitative or quantitative, the general consensus across studies was that pain was reported equivalently across assessment modalities. The meta-analysis of correlations between scores reported electronically and conventionally resulted in a summary coefficient of 0.92, indicating high correspondence. The summary coefficients produced by studies reporting ICC or weighted kappa and studies reporting Pearson or Spearman rho coefficients were not different from the overall summary score, suggesting negligible change in patient-reported scores across modalities. These findings agree with those of a meta-analysis published in 2008 that evaluated the equivalence of scores for patient-reported outcomes (not specifically pain) completed using PDA, computer, or tablet and paper-based methods and that showed a summary correlation of 0.90 [9]. Together, these reviews suggest that score equivalence between electronic and conventional data capture methods is a robust finding across patient-reported outcomes.

Despite our use of random effects models, we observed substantial heterogeneity across studies included in the meta-analysis that was not accounted for by the single study that met our criterion for extreme effect size, sensitivity analyses by correlation type, the similarity of pain assessment measure used in each modality, or duration of data collection. Studies varied in terms of study design, participant group, type of electronic and conventional data collection method, and pain measurement instrument—the heterogeneity may be explained by these differences in methodology. For instance, the type of electronic device used to collect pain data varied across studies, meaning that aspects of the device such as interface design, user familiarity, and screen size could each have contributed to our heterogeneous results [11]. The included studies also varied in terms of the type of pain intensity scale or pain interference instrument used (eg, NRS, VAS, etc). Although good congruence in patient self-report across instruments has been shown [75], and that the transfer of the assessment instrument to the electronic format generally appeared to be in good faith, as reported previously, differences in pain ratings across instruments are possible [76]. Irrespective of the observed heterogeneity, the correlation coefficients were strong across all studies with no reported coefficients less than 0.64, suggesting that heterogeneity should not temper the meta-analysis conclusion.

The collection of high-quality and complete patient-reported data is of utmost importance to clinicians, researchers, and study sponsors [12]. Data completeness was a commonly reported comparison outcome across data collection methods in the included studies. The results regarding the superiority of data completeness were mixed. However, the electronic method was most often associated with more complete data being collected. Ultimately, methodological and logistical issues related to paper-based data collection methods may support the use of electronic data capture. For instance, research has shown that the completeness and accuracy of pain data collected via paper methods is adversely impacted by patients back-filling diaries and, therefore, introducing recall bias into datasets (a behavior that can be rendered impossible using electronic methods) [8]. In addition, the capacity to efficiently and cost-effectively develop large databases for clinical and research purposes may be improved with electronic data capture. For instance, one of the studies included in this review [47] showed that over 4-fold more research assistant time was required to manage postoperative pain data collected using conventional means compared with electronic data. This finding suggests that cost savings may result from the use of electronic pain assessments in research, and this savings might be pronounced at scale. Furthermore, the likelihood of inaccurate or missing data in these databases resulting from human input error is reduced in the case of electronic entry [77].

Almost all studies that assessed ease indicated, in some manner, that electronic methods were easy to use, easy to understand, or easy to review or report pain. The time difference required to complete pain assessments via each data collection method was minimal, and the majority of studies showed that the electronic method required equal or less time to complete than conventional methods. The methodological advantages of electronic data capture include high-density sampling in all environments. Evidence of ease of use and efficiency in electronic data capture is useful to researchers and clinicians considering leveraging these utilities to collect repeated ecologically relevant pain assessments [78].

Electronic data capture was also shown to be a highly acceptable method for pain assessment and was more likely to be the method of choice for reporting by patients. These findings agree with those of previous studies comparing electronic and conventional methods [10]. Given the heterogeneity of electronic pain data capture methods, participant populations, and sampling densities of included studies, our results suggest acceptability across a range of data collection contexts. This result is meaningful as the acceptability of an intervention has been linked to adoption, especially in relation to long-term sustainability [79].

Limitations

Some included studies did not administer the same pain measurement instrument or use the same sampling schedule via electronic and conventional methods, making it difficult to directly compare results across modalities. Owing to variations in study design and the fact that our outcomes of interest were often times not the main objective of our included studies, we did not perform an assessment of quality for included studies; instead, we elected to include all identified studies in our review. Our results and conclusions are, therefore, the product of studies that may have included significant methodological weaknesses. In addition, as is an issue with all systematic reviews, we are constrained by possible publication bias, which was suggested by the funnel plot inspection of our quantitative synthesis data. However, given the objectives of the studies we included, we believe that the likelihood of a *file-drawer effect* is low. Finally, we included studies conducted in controlled (eg, research and health care institutions) and uncontrolled (eg, participant home) environments. We are, therefore, limited in our ability to make more definitive conclusions about our outcomes as they pertain to ecologically relevant data collection, which is considered a major methodological advantage of the electronic method.

Conclusions

Overall, this review demonstrates that electronic pain-related data capture methods are comparable with conventional methods in terms of score equivalence, data completeness, ease, efficiency, and acceptability. Specifically, pain-related outcome scores reported across methods were congruent in terms of score correlations and mean or median differences between scores. Data completeness, ease of use, efficiency, and acceptability outcomes were also comparable or superior using electronic data capture. Our results suggest that electronic methods are a feasible means to collect pain data, and the use of these methods is likely to increase with the ubiquitous use of mobile phones outside of the clinical or research setting. However, a critical caveat to this conclusion relates to the validation of pain instruments that are implemented electronically. To ensure the collection of accurate data, rigorous methods should be used to establish the sound psychometric properties of electronic pain measurement instruments. Validation of electronic methods will facilitate the capture of pain data in clinical settings but will also support its use in data collection for interventional research, an area that has largely not been explored to date [6].

Conflicts of Interest

PS works for and owns shares of a digital health company that makes electronic medical records. All other authors have no conflicts of interest to disclose.

Multimedia Appendix 1
Sample search strategy.
[DOCX File , 12 KB - jmir_v22i6e16480_app1.docx]

Multimedia Appendix 2
Funnel plot of 21 studies presenting correlations for score equivalence between electronic and conventional pain data collection modalities.
[PNG File , 48 KB - jmir_v22i6e16480_app2.png]

References

1. Melzack R. Evolution of the neuromatrix theory of pain. The Prithvi Raj lecture: presented at the third world congress of world institute of pain, Barcelona 2004. Pain Pract 2005 Jun;5(2):85-94. [doi: 10.1111/j.1533-2500.2005.05203.x] [Medline: 17177754]
2. Schiavenato M, Craig KD. Pain assessment as a social transaction: beyond the 'gold standard'. Clin J Pain 2010 Oct;26(8):667-676. [doi: 10.1097/AJP.0b013e3181e72507] [Medline: 20664341]
3. Asghari A, Nicholas MK. Pain self-efficacy beliefs and pain behaviour. A prospective study. Pain 2001 Oct;94(1):85-100. [doi: 10.1016/s0304-3959(01)00344-x] [Medline: 11576748]
4. Schneider S, Stone AA, Schwartz JE, Broderick JE. Peak and end effects in patients' daily recall of pain and fatigue: a within-subjects analysis. J Pain 2011 Mar;12(2):228-235 [FREE Full text] [doi: 10.1016/j.jpain.2010.07.001] [Medline: 20817615]
5. Stinson JN, Jibb LA, Lalloo C, Feldman BM, McGrath PJ, Petroz GC, et al. Comparison of average weekly pain using recalled paper and momentary assessment electronic diary reports in children with arthritis. Clin J Pain 2014 Dec;30(12):1044-1050. [doi: 10.1097/AJP.0000000000000072] [Medline: 24535055]

6. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. J Pain 2018 Jul;19(7):699-716 [FREE Full text] [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
7. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Construct validity and reliability of a real-time multidimensional smartphone app to assess pain in children and adolescents with cancer. Pain 2015 Dec;156(12):2607-2615. [doi: 10.1097/j.pain.0000000000000385] [Medline: 26580680]
8. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. Br Med J 2002 May 18;324(7347):1193-1194 [FREE Full text] [doi: 10.1136/bmj.324.7347.1193] [Medline: 12016186]
9. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
10. Lane SJ, Heddle NM, Arnold E, Walker I. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC Med Inform Decis Mak 2006 May 31;6:23 [FREE Full text] [doi: 10.1186/1472-6947-6-23] [Medline: 16737535]
11. Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27(7):MR000042. [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
12. Zbrozek A, Hebert J, Gogates G, Thorell R, Dell C, Molsen E, et al. Validation of electronic systems to collect patient-reported outcome (PRO) data-recommendations for clinical trial teams: report of the ISPOR ePRO systems validation good research practices task force. Value Health 2013 Jun;16(4):480-489 [FREE Full text] [doi: 10.1016/j.jval.2013.04.002] [Medline: 23796281]
13. von Korff M, Scher AI, Helmick C, Carter-Pokras O, Dodick DW, Goulet J, et al. United States national pain strategy for population research: concepts, definitions, and pilot data. J Pain 2016 Oct;17(10):1068-1080. [doi: 10.1016/j.jpain.2016.06.009] [Medline: 27377620]
14. Dillon DG, Pirie F, Rice S, Pomilla C, Sandhu MS, Motala AA, African Partnership for Chronic Disease Research (APCDR). Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa. J Clin Epidemiol 2014 Dec;67(12):1358-1363 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.06.012] [Medline: 25135245]
15. Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, Katz NP. Electronic diaries for monitoring chronic pain: 1-year validation study. Pain 2001 Apr;91(3):277-285. [doi: 10.1016/s0304-3959(00)00450-4] [Medline: 11275385]
16. Canadian Internet Registration Authority. 2018. Canada's Internet Factbook 2019 URL: https://www.cira.ca/resources/corporate/factbook/canadas-internet-factbook-2019 [accessed 2019-10-09]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
18. McGrath PJ, Walco GA, Turk DC, Dworkin RH, Brown MT, Davidson K, PedIMMPACT. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. J Pain 2008 Sep;9(9):771-783. [doi: 10.1016/j.jpain.2008.04.007] [Medline: 18562251]
19. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
21. Wood C, von Baeyer CL, Falinower S, Moyse D, Annequin D, Legout V. Electronic and paper versions of a faces pain intensity scale: concordance and preference in hospitalized children. BMC Pediatr 2011 Oct 12;11:87 [FREE Full text] [doi: 10.1186/1471-2431-11-87] [Medline: 21989306]
22. Castarlenas E, Sánchez-Rodríguez E, Vega RD, Roset R, Miró J. Agreement between verbal and electronic versions of the numerical rating scale (NRS-11) when used to assess pain intensity in adolescents. Clin J Pain 2015 Mar;31(3):229-234. [doi: 10.1097/AJP.0000000000000104] [Medline: 24699160]
23. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
24. Stinson JN, Stevens BJ, Feldman BM, Streiner D, McGrath PJ, Dupuis A, et al. Construct validity of a multidimensional electronic pain diary for adolescents with arthritis. Pain 2008 Jun;136(3):281-292. [doi: 10.1016/j.pain.2007.07.002] [Medline: 17723279]
25. Allena M, Cuzzoni MG, Tassorelli C, Nappi G, Antonaci F. An electronic diary on a palm device for headache monitoring: a preliminary experience. J Headache Pain 2012 Oct;13(7):537-541 [FREE Full text] [doi: 10.1007/s10194-012-0473-2] [Medline: 22842873]
26. Athale N, Sturley A, Skoczen S, Kavanaugh A, Lenert L. A web-compatible instrument for measuring self-reported disease activity in arthritis. J Rheumatol 2004 Mar;31(2):223-228. [Medline: 14760788]

27. Bandarian-Balooch S, Martin PR, McNally B, Brunelli A, Mackenzie S. Electronic-diary for recording headaches, triggers, and medication use: development and evaluation. Headache 2017 Nov;57(10):1551-1569. [doi: 10.1111/head.13184] [Medline: 28925498]
28. Bedson J, Hill J, White D, Chen Y, Wathall S, Dent S, et al. Development and validation of a pain monitoring app for patients with musculoskeletal conditions (the Keele pain recorder feasibility study). BMC Med Inform Decis Mak 2019 Jan 25;19(1):24 [FREE Full text] [doi: 10.1186/s12911-019-0741-z] [Medline: 30683106]
29. Bishop FL, Lewis G, Harris S, McKay N, Prentice P, Thiel H, et al. A within-subjects trial to test the equivalence of online and paper outcome measures: the Roland Morris disability questionnaire. BMC Musculoskelet Disord 2010 Jun 8;11:113 [FREE Full text] [doi: 10.1186/1471-2474-11-113] [Medline: 20529332]
30. Blum D, Koeberle D, Omlin A, Walker J, von Moos R, Mingrone W, et al. Feasibility and acceptance of electronic monitoring of symptoms and syndromes using a handheld computer in patients with advanced cancer in daily oncology practice. Support Care Cancer 2014 Sep;22(9):2425-2434. [doi: 10.1007/s00520-014-2201-8] [Medline: 24705855]
31. Byrom B, Doll H, Muehlhausen W, Flood E, Cassedy C, McDowell B, et al. Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. Value Health 2018 May;21(5):581-589 [FREE Full text] [doi: 10.1016/j.jval.2017.10.008] [Medline: 29753356]
32. Chiu LY, Sun T, Ree R, Dunsmuir D, Dotto A, Ansermino JM, et al. The evaluation of smartphone versions of the visual analogue scale and numeric rating scale as postoperative pain assessment tools: a prospective randomized trial. Can J Anaesth 2019 Jun;66(6):706-715. [doi: 10.1007/s12630-019-01324-9] [Medline: 30796700]
33. Christie A, Dagfinrud H, Dale Y, Schulz T, Hagen KB. Collection of patient-reported outcomes; text messages on mobile phones provide valid scores and high response rates. BMC Med Res Methodol 2014 Apr 16;14:52 [FREE Full text] [doi: 10.1186/1471-2288-14-52] [Medline: 24735061]
34. Cook AJ, Roberts DA, Henderson MD, van Winkle LC, Chastain DC, Hamill-Ruth RJ. Electronic pain questionnaires: a randomized, crossover comparison with paper questionnaires for chronic pain assessment. Pain 2004 Jul;110(1-2):310-317. [doi: 10.1016/j.pain.2004.04.012] [Medline: 15275781]
35. Cunha-Miranda L, Santos H, Miguel C, Silva C, Barcelos F, Borges J, et al. Validation of Portuguese-translated computer touch-screen questionnaires in patients with rheumatoid arthritis and spondyloarthritis, compared with paper formats. Rheumatol Int 2015 Dec;35(12):2029-2035. [doi: 10.1007/s00296-015-3347-5] [Medline: 26346588]
36. Fanciullo GJ, Cravero JP, Mudge BO, McHugo GJ, Baird JC. Development of a new computer method to assess children's pain. Pain Med 2007 Oct;8(Suppl 3):S121-S128. [doi: 10.1111/j.1526-4637.2007.00376.x] [Medline: 17877522]
37. Freynhagen R, Baron R, Tölle T, Stemmler E, Gockel U, Stevens M, et al. Screening of neuropathic pain components in patients with chronic back pain associated with nerve root compression: a prospective observational pilot study (MIPORT). Curr Med Res Opin 2006 Mar;22(3):529-537. [doi: 10.1185/030079906X89874] [Medline: 16574036]
38. Gaertner J, Elsner F, Pollmann-Dahmen K, Radbruch L, Sabatowski R. Electronic pain diary: a randomized crossover study. J Pain Symptom Manage 2004 Sep;28(3):259-267 [FREE Full text] [doi: 10.1016/j.jpainsymman.2003.12.017] [Medline: 15336338]
39. Garcia-Palacios A, Herrero R, Belmonte MA, Castilla D, Guixeres J, Molinari G, et al. Ecological momentary assessment for chronic pain in fibromyalgia using a smartphone: a randomized crossover study. Eur J Pain 2014 Jul;18(6):862-872. [doi: 10.1002/j.1532-2149.2013.00425.x] [Medline: 24921074]
40. Heiberg T, Kvien TK, Dale O, Mowinckel P, Aanerud GJ, Songe-Møller AB, et al. Daily health status registration (patient diary) in patients with rheumatoid arthritis: a comparison between personal digital assistant and paper-pencil format. Arthritis Rheum 2007 Apr 15;57(3):454-460 [FREE Full text] [doi: 10.1002/art.22613] [Medline: 17394232]
41. Hofstedt O, di Giuseppe D, Alenius G, Stattin N, Forsblad-D'Elia H, Ljung L. Comparison of agreement between internet-based registration of patient-reported outcomes and clinic-based paper forms within the Swedish rheumatology quality register. Scand J Rheumatol 2019 Jul;48(4):326-330. [doi: 10.1080/03009742.2018.1551964] [Medline: 30758242]
42. Jaatun EA, Hjermstad MJ, Gundersen OE, Oldervoll L, Kaasa S, Haugen DF, European Palliative Care Research Collaborative (EPCRC). Development and testing of a computerized pain body map in patients with advanced cancer. J Pain Symptom Manage 2014 Jan;47(1):45-56 [FREE Full text] [doi: 10.1016/j.jpainsymman.2013.02.025] [Medline: 23856098]
43. Jamison RN, Gracely RH, Raymond SA, Levine JG, Marino B, Herrmann TJ, et al. Comparative study of electronic vs paper VAS ratings: a randomized, crossover trial using healthy volunteers. Pain 2002 Sep;99(1-2):341-347. [doi: 10.1016/s0304-3959(02)00178-1] [Medline: 12237213]
44. Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. J Pain 2006 Mar;7(3):192-199. [doi: 10.1016/j.jpain.2005.10.006] [Medline: 16516825]
45. Jonassaint CR, Shah N, Jonassaint J, de Castro L. Usability and feasibility of an mhealth intervention for monitoring and managing pain symptoms in sickle cell disease: the sickle cell disease mobile application to record symptoms via technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]

46. Junker U, Freynhagen R, Längler K, Gockel U, Schmidt U, Tölle TR, et al. Paper versus electronic rating scales for pain assessment: a prospective, randomised, cross-over validation study with 200 chronic pain patients. Curr Med Res Opin 2008 Jun;24(6):1797-1806. [doi: 10.1185/03007990802121059] [Medline: 18485269]
47. Khan J, Jibb L, Busse J, Gilron I, Choi S, Paul J, et al. Electronic versus traditional data collection: a multicenter randomized controlled perioperative pain trial. Canadian J Pain 2019 Jul 30;3(2):16-25. [doi: 10.1080/24740527.2019.1587584]
48. Kim CH, Chung CK, Choi Y, Shin H, Woo JW, Kim S, et al. The usefulness of a mobile device-based system for patient-reported outcomes in a spine outpatient clinic. Spine J 2016 Jul;16(7):843-850. [doi: 10.1016/j.spinee.2016.02.048] [Medline: 26961200]
49. Koho P, Aho S, Kautiainen H, Pohjolainen T, Hurri H. Test-retest reliability and comparability of paper and computer questionnaires for the Finnish version of the Tampa Scale of Kinesiophobia. Physiotherapy 2014 Dec;100(4):356-362. [doi: 10.1016/j.physio.2013.11.007] [Medline: 24679373]
50. Kvien TK, Mowinckel P, Heiberg T, Dammann KL, Dale O, Aanerud GJ, et al. Performance of health status measures with a pen based personal digital assistant. Ann Rheum Dis 2005 Oct;64(10):1480-1484 [FREE Full text] [doi: 10.1136/ard.2004.030437] [Medline: 15843456]
51. MacKenzie H, Thavaneswaran A, Chandran V, Gladman DD. Patient-reported outcome in psoriatic arthritis: a comparison of web-based versus paper-completed questionnaires. J Rheumatol 2011 Dec;38(12):2619-2624. [doi: 10.3899/jrheum.110165] [Medline: 22045844]
52. Marceau LD, Link C, Jamison RN, Carolan S. Electronic diaries as a tool to improve pain management: is there any evidence? Pain Med 2007 Oct;8(Suppl 3):S101-S109. [doi: 10.1111/j.1526-4637.2007.00374.x] [Medline: 17877520]
53. Marceau LD, Link CL, Smith LD, Carolan SJ, Jamison RN. In-clinic use of electronic pain diaries: barriers of implementation among pain physicians. J Pain Symptom Manage 2010 Sep;40(3):391-404 [FREE Full text] [doi: 10.1016/j.jpainsymman.2009.12.021] [Medline: 20580526]
54. Matthews M, Rathleff MS, Vicenzino B, Boudreau SA. Capturing patient-reported area of knee pain: a concurrent validity study using digital technology in patients with patellofemoral pain. PeerJ 2018;6:e4406 [FREE Full text] [doi: 10.7717/peerj.4406] [Medline: 29568700]
55. Neudecker J, Raue W, Schwenk W. High correlation but inadequate point-to-point agreement, between conventional mechanical and electronical visual analogue scale for assessment of acute postoperative pain after general surgery. Acute Pain 2006 Dec;8(4):175-180. [doi: 10.1016/j.acpain.2006.08.043]
56. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004 Mar;107(3):213-219. [doi: 10.1016/j.pain.2003.10.005] [Medline: 14736583]
57. Pawar SG, Ramani PS, Prasad A, Dhar A, Babhulkar SS, Bahurupi YA. Software version of Roland Morris disability questionnaire for outcome assessment in low back pain. Neurol Res 2017 Apr;39(4):292-297. [doi: 10.1080/01616412.2017.1297555] [Medline: 28337948]
58. Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. J Med Internet Res 2004 Sep 15;6(3):e29 [FREE Full text] [doi: 10.2196/jmir.6.3.e29] [Medline: 15471755]
59. Rolfson O, Salomonsson R, Dahlberg LE, Garellick G. Internet-based follow-up questionnaire for measuring patient-reported outcome after total hip replacement surgery-reliability and response rate. Value Health 2011;14(2):316-321 [FREE Full text] [doi: 10.1016/j.jval.2010.08.004] [Medline: 21402299]
60. Saleh KJ, Radosevich DM, Kassim RA, Moussa M, Dykes D, Bottolfson H, et al. Comparison of commonly used orthopaedic outcome measures using palm-top computers and paper surveys. J Orthop Res 2002 Nov;20(6):1146-1151 [FREE Full text] [doi: 10.1016/S0736-0266(02)00059-1] [Medline: 12472221]
61. Sánchez-Rodríguez E, de la Vega R, Castarlenas E, Roset R, Miró J. An app for the assessment of pain intensity: validity properties and agreement of pain reports when used with young people. Pain Med 2015 Oct;16(10):1982-1992. [doi: 10.1111/pme.12859] [Medline: 26176297]
62. Serif T, Ghinea G. Recording of time-varying back-pain data: a wireless solution. IEEE Trans Inf Technol Biomed 2005 Sep;9(3):447-458. [doi: 10.1109/titb.2005.847514] [Medline: 16167699]
63. Stinson JN, Connelly M, Jibb LA, Schanberg LE, Walco G, Spiegel LR, et al. Developing a standardized approach to the assessment of pain in children and youth presenting to pediatric rheumatology providers: a Delphi survey and consensus conference process followed by feasibility testing. Pediatr Rheumatol Online J 2012 Apr 10;10(1):7 [FREE Full text] [doi: 10.1186/1546-0096-10-7] [Medline: 22490427]
64. Stomberg MW, Platon B, Widén A, Wallner I, Karlsson O. Health information: what can mobile phone assessments add? Perspect Health Inf Manag 2012;9:1-10 [FREE Full text] [Medline: 23209453]
65. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain 2003 Jul;104(1-2):343-351. [doi: 10.1016/s0304-3959(03)00040-x] [Medline: 12855344]
66. Sun T, West N, Ansermino JM, Montgomery CJ, Myers D, Dunsmuir D, et al. A smartphone version of the faces pain scale-revised and the color analog scale for postoperative pain assessment in children. Paediatr Anaesth 2015 Dec;25(12):1264-1273. [doi: 10.1111/pan.12790] [Medline: 26507916]

67. Suso-Ribera C, Castilla D, Zaragozá I, Ribera-Canudas MV, Botella C, García-Palacios A. Validity, reliability, feasibility, and usefulness of pain monitor: a multidimensional smartphone app for daily monitoring of adults with heterogenous chronic pain. Clin J Pain 2018 Oct;34(10):900-908. [doi: 10.1097/AJP.0000000000000618] [Medline: 29659375]
68. Symonds T, Hughes B, Liao S, Ang Q, Bellamy N. Validation of the Chinese western Ontario and McMaster universities osteoarthritis index in patients from mainland china with osteoarthritis of the knee. Arthritis Care Res (Hoboken) 2015 Nov;67(11):1553-1560 [FREE Full text] [doi: 10.1002/acr.22631] [Medline: 26018634]
69. Theiler R, Alon E, Brugger S, Ljutow A, Mietzsch T, Müller D, et al. Evaluation of a standardized internet-based and telephone-based patient monitoring system for pain therapy with transdermal fentanyl. Clin J Pain 2007;23(9):804-811. [doi: 10.1097/AJP.0b013e3181565d04] [Medline: 18075409]
70. van den Kerkhof EG, Goldstein DH, Lane J, Rimmer MJ, van Dijk JP. Using a personal digital assistant enhances gathering of patient data on an acute pain management service: a pilot study. Can J Anaesth 2003 Apr;50(4):368-375. [doi: 10.1007/BF03021034] [Medline: 12670814]
71. van den Kerkhof EG, Goldstein DH, Rimmer MJ, Tod DA, Lee HK. Evaluation of hand-held computers compared to pen and paper for documentation on an acute pain service. Acute Pain 2004 Dec;6(3-4):115-121. [doi: 10.1016/j.acpain.2004.07.001]
72. Wæhrens EE, Amris K, Bartels EM, Christensen R, Danneskiold-Samsøe B, Bliddal H, et al. Agreement between touch-screen and paper-based patient-reported outcomes for patients with fibromyalgia: a randomized cross-over reproducibility study. Scand J Rheumatol 2015;44(6):503-510. [doi: 10.3109/03009742.2015.1029517] [Medline: 26114582]
73. Sanchez MA, Rabin BA, Gaglio B, Henton M, Elzarrad MK, Purcell P, et al. A systematic review of eHealth cancer prevention and control interventions: new technology, same methods and designs? Transl Behav Med 2013 Dec;3(4):392-401 [FREE Full text] [doi: 10.1007/s13142-013-0224-1] [Medline: 24294327]
74. Steele L, Lade H, McKenzie S, Russell TG. Assessment and diagnosis of musculoskeletal shoulder disorders over the internet. Int J Telemed Appl 2012;2012:945745 [FREE Full text] [doi: 10.1155/2012/945745] [Medline: 23193395]
75. Connelly M, Neville K. Comparative prospective evaluation of the responsiveness of single-item pediatric pain-intensity self-report scales and their uniqueness from negative affect in a hospital setting. J Pain 2010 Dec;11(12):1451-1460. [doi: 10.1016/j.jpain.2010.04.011] [Medline: 20646966]
76. Gagliese L, Katz J. Age differences in postoperative pain are scale dependent: a comparison of measures of pain intensity and quality in younger and older surgical patients. Pain 2003 May;103(1-2):11-20. [doi: 10.1016/s0304-3959(02)00327-5] [Medline: 12749954]
77. Fleischmann R, Decker A, Kraft A, Mai K, Schmidt S. Mobile electronic versus paper case report forms in clinical trials: a randomized controlled trial. BMC Med Res Methodol 2017 Dec 1;17(1):153 [FREE Full text] [doi: 10.1186/s12874-017-0429-y] [Medline: 29191176]
78. Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. PLoS One 2011;6(9):e25348 [FREE Full text] [doi: 10.1371/journal.pone.0025348] [Medline: 21966505]
79. Walker I, Sigouin C, Sek J, Almonte T, Carruthers J, Chan A, et al. Comparing hand-held computers and paper diaries for haemophilia home therapy: a randomized trial. Haemophilia 2004 Nov;10(6):698-704. [doi: 10.1111/j.1365-2516.2004.01046.x] [Medline: 15569164]

Abbreviations

ICC: intraclass correlation coefficient
NRS: Numerical Rating Scale
PDA: personal digital assistant
VAS: Visual Analog Scale

Edited by G Eysenbach; submitted 16.10.19; peer-reviewed by S Haghayegh, E Sadeghi-Demneh; comments to author 14.11.19; revised version received 21.01.20; accepted 22.03.20; published 16.06.20.

Please cite as:
Jibb LA, Khan JS, Seth P, Lalloo C, Mulrooney L, Nicholson K, Nowak DA, Kaur H, Chee-A-Tow A, Foster J, Stinson JN
Electronic Data Capture Versus Conventional Data Collection Methods in Clinical Pain Studies: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(6):e16480
URL: https://www.jmir.org/2020/6/e16480
doi:10.2196/16480
PMID:32348259

Original Paper

Attitudes and Use of Information and Communication Technologies in Older Adults With Mild Cognitive Impairment or Early Stages of Dementia and Their Caregivers: Cross-Sectional Study

Jose Guzman-Parra[1*], PhD; Pilar Barnestein-Fonseca[1*], PhD; Gloria Guerrero-Pertiñez[1], PhD; Peter Anderberg[2], PhD; Luis Jimenez-Fernandez[1], BE; Esperanza Valero-Moreno[1], MA; Jessica Marian Goodman-Casanova[1], RNCS; Antonio Cuesta-Vargas[3], PhD; Maite Garolera[4], PhD; Maria Quintana[4], PhD; Rebeca I García-Betances[5], PhD; Evi Lemmens[6], PhD; Johan Sanmartin Berglund[2], MD, PhD; Fermin Mayoral-Cleries[1], MD, PhD

[1]Mental Health Department, Instituto de Investigación Biomédica de Málaga, University Regional Hospital of Malaga, Malaga, Spain
[2]Department of Health, Blekinge Institute of Technology, Karlskrona, Sweden
[3]Departamento de Fisioterapia, Instituto de Biomedicina de Málaga, Universidad de Málaga, Malaga, Spain
[4]Brain, Cognition and Behavior – Clinical Research, Consorci Sanitari de Terrassa, Barcelona, Spain
[5]Life Supporting Technologies Group, Universidad Politécnica, Madrid, Spain
[6]University Colleges Leuven-Limburg, Genk, Belgium
[*]these authors contributed equally

Corresponding Author:
Jose Guzman-Parra, PhD
Mental Health Department
Instituto de Investigación Biomédica de Málaga
University Regional Hospital of Malaga
Plaza Hospital Civil s/n Hospital Civil 1ª Planta Pabellón 4
Malaga
Spain
Phone: 34 951 29 03 07
Email: jgp00004@hotmail.com

Abstract

Background: Information and communication technologies are promising tools to increase the quality of life of people with dementia or mild cognitive impairment and that of their caregivers. However, there are barriers to their use associated with sociodemographic factors and negative attitudes, as well as inadequate knowledge about technologies.

Objective: The aim of this study was to analyze technophilia (attitudes toward new technologies) and the use of smartphones and tablets along with associated factors in people with dementia/mild cognitive impairment and their caregivers.

Methods: Data from the first visit of the Support Monitoring and Reminder for Mild Dementia (SMART4MD) randomized multicenter clinical trial were used for this analysis. Data were obtained from two European countries, Spain and Sweden, and from three centers: Consorci Sanitari de Terrassa (Catalonia, Spain), Servicio Andaluz de Salud (Andalusia, Spain), and the Blekinge Institute of Technology (Sweden). Participants with a score between 20 and 28 in the Mini Mental State Examination, with memory problems (for more than 6 months), and who were over the age of 55 years were included in the study, along with their caregivers. The bivariate Chi square and Mann-Whitney tests, and multivariate linear and logistic regression models were used for statistical analysis.

Results: A total of 1086 dyads were included (N=2172). Overall, 299 (27.53%) of people with dementia/mild cognitive impairment had a diagnosis of dementia. In addition, 588 (54.14%) of people with dementia/mild cognitive impairment reported using a smartphone almost every day, and 106 (9.76%) used specific apps or software to support their memory. Among the caregivers, 839 (77.26%) used smartphones and tablets almost every day, and 181 (16.67%) used specific apps or software to support their memory. The people with dementia/mild cognitive impairment showed a lower level of technophilia in comparison to that of their caregivers after adjusting for confounders (B=0.074, *P*=.02) with differences in technology enthusiasm (B=0.360, *P*<.001), but not in technology anxiety (B=–0.042, *P*=.37). Technophilia was associated with lower age (B=–0.009, *P*=.004), male gender (B=–0.160, *P*<.001), higher education level (*P*=.01), living arrangement (living with children vs single; B=–2.538, *P*=.01), country of residence (Sweden vs Spain; B=0.256, *P*<.001), lower depression (B=–0.046, *P*<.001), and better health status (B=0.004, *P*<.001) in people with dementia/mild cognitive impairment. Among caregivers, technophilia was associated with comparable sociodemographic factors (except for living arrangement), along with a lower caregiver burden (B=–0.005, *P*=.04) and better quality of life (B=0.348, *P*<.001).

Conclusions: Technophilia was associated with a better quality of life and sociodemographic variables in people with dementia/mild cognitive impairment and caregivers, suggesting potential barriers for technological interventions. People with dementia/mild cognitive impairment frequently use smartphones and tablets, but the use of specific apps or software to support memory is limited. Interventions using these technologies are needed to overcome barriers in this population related to sociodemographic characteristics and the lack of enthusiasm for new technologies.

Trial Registration: ClinicalTrials.gov NCT03325699; https://clinicaltrials.gov/ct2/show/NCT03325699

(J Med Internet Res 2020;22(6):e17253) doi:10.2196/17253

KEYWORDS

aging; mild cognitive impairment; dementia eHealth; information and communication technology; technophilia

Introduction

The population in Europe is getting older, and consequently the number of people with dementia or mild cognitive impairment is increasing given the association of these conditions with age. It is estimated that the population with dementia will double by 2030 and triple by 2050, reaching more than 115 million individuals [1]. One of the core symptoms in mild cognitive impairment and in the early stages of dementia is memory impairment, which is a condition that is also associated with depression, sleep problems, and other behavioral symptoms [2,3]. Guaranteeing an optimal quality of life for people with these conditions remains an enormous challenge because there are no effective long-term pharmacological treatments in the majority of cases [4].

Information and communication technologies (ICTs), especially touchscreen technologies, are promising tools to increase the quality of life and cognitive function of people with dementia/mild cognitive impairment and their caregivers [5-9]. These technologies could be used to train cognitive functions, monitor health and movements, provide reminders to support memory, promote social support, improve communication with caregivers, and provide useful information about the condition. Smartphones and tablets have the advantage of not raising a stigma for the individuals that use them [10] because they are ubiquitous and used by the majority of the population, and they also represent a natural source of data for professionals and researchers [10-13]. Over 40,000 health-related apps exist but very few are specifically designed for people with dementia/mild cognitive impairment [14]. Previous studies have found barriers associated with the use of ICTs in older adults, including age and education level [15]. Likewise, barriers for the use of technology in people with dementia/mild cognitive impairment have been described, including negative attitudes toward ICTs, inaccurate perceptions of ICTs, and poor technology knowledge [16]. Although the attitudes and knowledge of ICTs and touchscreen technologies could determine the use of health apps, to the best of our knowledge, few large-sample studies have investigated the attitude and use of these technologies in older adults, including people with dementia/mild cognitive impairment and their caregivers. Technophilia is one of the emerging concepts regarding attitudes toward technologies. One of the definitions of technophilia is the "attraction, enthusiasm of the human individual determined by the activities which involve the use of advanced technologies. It is expressed by easy adaptation to the social changes brought by technological innovations" [17].

The aim of this study was to analyze technophilia in people with dementia/mild cognitive impairment and their caregivers and to determine the sociodemographic and clinical factors associated with technophilia. Another aim was to analyze how this population uses smartphones, tablets, apps, and software to support their memory, and to identify factors associated with the use of apps to support memory. We tested the hypothesis that different sociodemographic and clinical factors in people with dementia/mild cognitive impairment and their caregivers are associated with technophilia and the use of specific apps to support memory in tablets or smartphones.

Methods

Study Design

In this cross-sectional study, data from the baseline assessment of the Support Monitoring and Reminder for Mild Dementia (SMART4MD) randomized multicenter clinical trial (ClinicalTrials.gov NCT03325699) were used. The objective of the trial was to create a digital platform (SMART4MD) for a tablet and to test if the platform had an impact on the quality of life of people with dementia/mild cognitive impairment and their caregivers. More detailed information on the trial is available in the published protocol [18].

Setting

This study was carried out in two European countries, Spain and Sweden, and at three centers: Consorci Sanitari de Terrassa (Catalonia, Spain), Servicio Andaluz de Salud (Andalusia, Spain), and the Blekinge Institute of Technology (Sweden).

Participants

A total of 1086 participant dyads (N=2172) were included in the study. The participant dyads comprised people with dementia/mild cognitive impairment and their informal caregivers. The participants were selected using a nonprobabilistic consecutive sampling method. The inclusion criteria were as follows: (1) score of 20 to 28 points on the Mini Mental State Examination (MMSE), (2) experience of memory problems over a substantial period of time (more than 6 months), (3) aged>55 years, (4) recipients of home care, (5) have an informal caregiver, (6) taking prescribed medication and in charge of it, and (7) no physical conditions that would reduce their ability to use a touchscreen app. The exclusion criteria were as follows: (1) terminal illness with less than 3 years of expected survival, (2) score above 11 on the Geriatric Depression Scale (GDS-15), or (3) another known significant cause of disease as an explanation for cognitive impairment such as substance abuse, bipolar disorder, schizophrenia, or developmental disorders.

Measures

Dependent Variables

The dependent variables were technophilia and use and familiarity with touchscreen devices. The TechPH questionnaire was used to assess technophilia [19]. This questionnaire includes 6 items assessed on a 5-point Likert scale from 1 (fully disagree) to 5 (fully agree), which was designed to specifically assess technophilia in the older population. The scale has two factors: technology enthusiasm and technology anxiety. The TechPH index is a score derived from the 6 items (the sum of items divided by 6) ranging from 1 to 5. More information on the TechPH index is included in the report on the validation study [19].

Use and familiarity with touchscreen devices was assessed with a questionnaire tailored to this study covering the following aspects: (1) use of a smartphone and tablets ("On average how often would you say you have been using a smartphone or tablet during the last 3 months?"), (2) use of the internet ("How often do you use the internet on your smartphone or tablet?"), (3) knowledge ("How knowledgeable do you consider yourself when it comes to using a smartphone or a tablet?"), (4) use of technology to support memory ("Are you using your mobile phone or tablet as a way to support your memory today?"), (5) use of specific apps for memory ("Do you have any special app or software on your mobile phone or tablet that you use to support your memory?"), and (6) perspective on the helpfulness of the technology for their memory ("Do you think that using your mobile phone or tablet to support your memory helps you to remember things?").

Independent Variables in People With Dementia/Mild Cognitive Impairment

Health-related quality of life was measured using the total score of the Quality of Life in Alzheimer's Disease (QoL-AD) questionnaire [20], which is a 13-item measure with a 4-point Likert scale. The EuroQoL-5D-3L [21] (EQ5D) questionnaire was also administered, which is a self-completion questionnaire that consists of 5 questions along with a scale for the participant to rate their health state on a scale thermometer of 0 to 100 (EQ-VAS). The European value set of Köning et al [22] was used for calculation of the EQ5D score.

Functional decline was assessed with the Lawton Instrumental Activities of Daily Living (IADL) [23] scale in people with dementia/mild cognitive impairment. The IADL uses 8 items for women and 5 items for men. The score was rescaled to range between 0 and 1.

The severity of cognitive impairment was assessed with the MMSE [24]. Depression was scored with the GDS-15 [25], which is a widely used scale to assess geriatric depression with 15 items and a range of 0-4 (normal), 5-8 (mildly depressed), 9-11 (moderately depressed), and 12-15 (severely depressed). Since an inclusion criterion for study participation was a GDS-15 score<11, the range in this study was 0–11.

Sociodemographic data included age, gender, education, living arrangement, marital status, and country of residence. Medical history included a diagnosis of dementia and comorbidity based on the International Classification of Diseases-10 [26].

Independent Variables for Caregivers

The independent variables for caregivers included the caregiver burden, which was assessed using the Zarit Caregiver Burden Interview (ZBI-12) [27,28], a 12-item 5-point Likert scale questionnaire. This is the short version of the original scale and was designed specifically for caregivers of people with dementia/mild cognitive impairment. Health-related quality of life was measured using the EQ5D as described above. The sociodemographic data included age, gender, education, living arrangements, marital status, country of residence, and relationship with the patient.

Statistical Analysis

The mean (SD) and frequency and percentages are used to describe continuous and categorical variables, respectively. To compare groups with high and low technophilia, we used the median TechPh Index (2.83 in people with dementia/mild cognitive impairment and 3.00 in caregivers) as the cut-off point. To compare groups, we used the Chi square test for categorical variables and the Mann-Whitney test for continuous variables. A multivariate linear regression analysis was conducted to analyze the factors associated with technophilia. To check the assumptions of the linear model, the Breusch-Pagan test (homoscedasticity), Shapiro-Wilk normality test of the residuals, variance inflation factor (VIF; multicollinearity), and a scatter plot for the linearity of the variables were used. We also analyzed the factors associated with the use of specific apps or software to support memory using multivariate logistic regression. To check the goodness of fit, we used the Hosmer-Lemeshow test and the VIF for multicollinearity. If the VIF was higher than 2, the variable was taken out of the model. A 95% level of significance was used for assessment. R program version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) and the R Commander package were used for these analyses.

Results

The basic characteristics of the sample of people with dementia/mild cognitive impairment are summarized in Table 1. Nearly a third of the sample were diagnosed with dementia. A total of 588 of the 1086 people with dementia/mild cognitive impairment (54.14%) used smartphones and tablets almost every day and 284 (26.15%) had never used these technologies. Only 381/1086 (35.08%) used the internet on smartphones and tablets almost every day and 470/1086 (43.28%) never used the internet on these gadgets. More than half (706/1086, 65.01%) considered themselves not at all or quite knowledgeable when it comes to using a smartphone or a tablet. A total of 106 (9.76%) had a special app or software on their mobile phone or tablet to support memory, and 669/1086 (61.60%) believed that using a mobile phone or tablet to support their memory helps them remember things. Additional information about the use of smartphones and tablets by the people with dementia/mild cognitive impairment is provided in Multimedia Appendix 1. Information about groups with high and low technophilia and groups that use or do not use specific apps or software on their mobile phone or tablet to support memory is shown in Table 1.

Table 1. Characteristics of participants with dementia or mild cognitive impairment.

Variables	All participants (N=1086)[a]	High technophilia (N=591)[b]	Low technophilia (N=493)[c]	*P* value	Use apps to support memory (N=106)[d]	Do not use apps to support memory (N=877)[e]	*P* value
Age (years), mean (SD)	74.48 (7.24)	74.32 (7.32)	74.57 (7.16)	.72	70.22 (7.46)	74.76 (7.11)	<.001
Gender, n (%)				<.001			.29
Female	576 (53.0)	276 (46.7)	298 (60.4)		51 (48.1)	470 (53.6)	
Male	510 (47.0)	315 (53.3)	195 (39.6)		55 (51.9)	407 (46.4)	
Education level, n (%)				.003			.03
Elementary school	653 (60.4)	331 (56.3)	320 (65.2)		55 (51.9)	527 (60.4)	
Secondary school	225 (20.8)	126 (21.4)	99 (20.2)		21 (19.8)	190 (21.8)	
Higher education	203 (18.8)	131 (22.2)	72 (14.7)		30 (28.3)	155 (17.8)	
Marital status, n (%)				.06			.06
Unmarried	41 (3.8)	20 (3.4)	21 (4.3)		6 (5.7)	31 (3.5)	
Married	697 (64.2)	377 (43.9)	320 (64.9)		75 (70.8)	555 (63.4)	
Common law partner	44 (4.1)	33 (5.6)	11 (2.2)		7 (6.6)	35 (4.0)	
Divorced	58 (5.3)	34 (5.8)	24 (4.9)		5 (4.7)	49 (5.6)	
Widowed	245 (22.6)	126 (21.4)	117 (23.7)		13 (12.3)	206 (23.5)	
Living arrangement, n (%)				.22			.05
Single	222 (20.6)	109 (18.6)	113 (23.1)		19 (17.9)	177 (20.4)	
Spouse/partner	686 (63.6)	385 (65.7)	301 (61.6)		73 (68.9)	549 (63.2)	
Children	97 (9.0)	57 (9.7)	40 (8.2)		3 (2.8)	85 (9.8)	
Other	73 (6.8)	36 (6.1)	35 (7.2)		11 (10.4)	58 (6.7)	
Country of residence, n (%)				<.001			.12
Sweden	345 (31.8)	231 (39.1)	114 (23.1)		39 (36.8)	258 (29.4)	
Spain	741 (68.2)	360 (60.9)	379 (76.9)		67 (63.2)	619 (70.6)	
Diagnosis of dementia, n (%)				.73			.007
Yes	299 (28.5)	165 (28.8)	132 (27.8)		87 (82.9)	594 (70.2)	
No	750 (71.5)	408 (71.2)	342 (72.2)		18 (17.1)	252 (29.8)	
Diagnosis of other medical condition, n (%)				.04			.96
Yes	267 (24.6)	431 (72.9)	386 (78.3)		26 (24.5)	213 (24.3)	
No	819 (75.4)	160 (27.1)	107 (21.7)		80 (75.5)	664 (75.7)	
Cognitive Status (MMSE[f]), mean (SD)	25.41 (2.48)	25.51 (2.53)	25.32 (2.40)	.05	25.97 (2.05)	25.30 (2.56)	.03
Depression (GDS-15[g]), mean (SD)	3.03 (2.84)	2.47 (2.53)	3.70 (3.05)	<.001	3.06 (3.01)	3.11 (2.84)	.53
Instrumental activities (IADL[h]), mean (SD)	0.87 (0.19)	0.87 (0.20)	0.86 (0.19)	.07	0.91 (0.17)	0.86 (0.19)	.008
Quality of life (QoL-AD[i]), mean (SD)	36.08 (6.52)	37.67 (6.31)	34.21 (6.27)	<.001	36.93 (7.32)	36.00 (6.49)	.21
Quality of life (EQ5D[j]), mean (SD)	0.75 (0.22)	0.78 (0.22)	0.72 (0.22)	<.001	0.69 (0.25)	0.75 (0.22)	.08
Health State thermometer, mean (SD)	69.45 (19.60)	72.98 (18.70)	65.27 (19.88)	<.001	70.67 (19.03)	69.22 (19.80)	.43

[a] Education level N=1081; Living arrangement N=1078; Marital status N=1085; Diagnosis of dementia N=1049.

[b] Education level N=588; Marital status N=590; Living arrangement N=586; Diagnosis of dementia N=573.

[c] Education level N=491; Living arrangement N=489; Diagnosis of dementia N=474.

[d] Diagnosis of dementia N=105.

[e] Education level N=872; Marital status N=876; Living arrangement N=869; Diagnosis of dementia N=846.

[f] MMSE: Mini Mental State Examination.

[g] GDS-15: Geriatric Depression Scale.

[h] IADL: Lawton Instrumental Activities of Daily Living.

[i] QoL-AD: Quality of Life in Alzheimer's Disease.

[j] EQ5D: EuroQoL-5D-3L.

The basic characteristics of the caregivers are summarized in Table 2. Among the 1086 caregivers, 839 (77.26%) used touchscreen technologies almost every day, and 123 (11.33%) had never used smartphones or tablets. A total of 721 (66.39%) used the internet on smartphones and tablets almost every day, and 191 (17.59%) never used the internet on touchscreen gadgets. A total of 433 (39.87%) considered themselves not at all or quite knowledgeable when it comes to using a smartphone or a tablet. A total of 181 (16.67%) had a special app or software on their mobile phone or tablet to support their memory, and 773 (71.18%) believed that using a mobile phone or tablet to support their memory helps them to remember things. Additional information about the use of smartphones and tablets by the caregivers is provided in Multimedia Appendix 1. Information about the caregivers overall, groups with high and low technophilia, and groups that use or do not use specific apps or software on their tablet or mobile phone to support memory is shown in Table 2.

Table 2. Characteristics of caregivers.

Variables	Total (N=1086)[a]	High technophilia (N=464)[b]	Low technophilia (N=598)[c]	*P* value	Use apps to support memory (N=181)[d]	Do not use apps to support memory (N=840)[e]	*P* value
Age (years), mean (SD)	62.29 (14.68)	58.35 (15.84)	65.24 (12.95)	<.001	54.27 (14.49)	63.34 (14.20)	<.001
Gender, n (%)				.003			.11
Female	741 (68.2)	295 (63.6)	167 (27.9)		133 (73.5)	566 (67.4)	
Male	345 (31.8)	169 (36.4)	431 (72.1)		48 (26.5)	274 (32.6)	
Education level, n (%)				<.001			<.001
Elementary school	378 (35.7)	111 (24.3)	257 (44.2)		29 (16.2)	318 (38.8)	
Secondary school	345 (32.5)	167 (36.6)	173 (29.3)		73 (40.8)	256 (31.3)	
Higher education	337 (31.8)	178 (39.0)	152 (26.1)		77 (43.0)	245 (29.9)	
Marital status, n (%)				<.001			.005
Unmarried	126 (11.6)	75 (16.2)	46 (7.7)		33 (18.2)	88 (10.5)	
Married	804 (74.2)	312 (67.4)	475 (79.6)		121 (66.9)	633 (75.5)	
Common law partner	76 (7.0)	38 (8.2)	37 (6.2)		15 (8.3)	55 (6.6)	
Divorced	47 (4.3)	23 (5.0)	24 (4.0)		11 (6.1)	34 (4.1)	
Widowed	30 (2.8)	15 (3.2)	15 (2.5)		1 (0.5)	28 (3.3)	
Living arrangement, n (%)				.005			.30
Single	98 (9.0)	50 (10.8)	45 (7.5)		17 (9.4)	75 (8.9)	
Spouse/partner	738 (68.1)	291 (62.9)	436 (73.0)		112 (61.9)	576 (68.7)	
Children	100 (9.2)	52 (11.2)	45 (7.5)		21 (11.6)	77 (9.2)	
Other	147 (13.6)	70 (15.1)	71 (11.9)		31 (17.1)	110 (13.1)	
Country of residence, n (%)				.03			.05
Sweden	345 (31.8)	167 (36.0)	177 (29.6)		44 (24.3)	266 (31.7)	
Spain	741 (68.2)	297 (64.0)	421 (70.4)		137 (75.7)	574 (68.3)	
Relation with the patient, n (%)				<.001			<.001
Spouse/partner	590 (55.5)	204 (44.7)	374 (63.9)		59 (33.7)	483 (58.6)	
Child	320 (30.1)	179 (39.3)	134 (22.9)		97 (55.4)	214 (26.0)	
Other	153 (14.4)	73 (16.0)	77 (13.2)		19 (10.9)	127 (15.4)	
Caregiver Burden (ZBI-12[f]), mean (SD)	6.86 (7.70)	6.31 (7.21)	7.28 (8.08)	.13	9.02 (7.62)	6.53 (7.68)	<.001
Quality of life (EQ5D[g]), mean (SD)	0.77 (0.21)	0.81 (0.20)	0.75 (0.22)	<.001	0.77 (0.22)	0.78 (0.21)	.72
Health State (thermometer), mean (SD)	72.24 (18.62)	74.95 (17.36)	70.07 (19.22)	<.001	70.94 (18.21)	72.46 (18.67)	.22

[a] Education level N=1060; Marital status N=1083; Living arrangement N=1083; Relation with the patient N=1063.

[b] Education level N=456; Marital status N=463; Living arrangement N=463; Relation with the patient N=456.

[c] Education level N=582; Marital status N=597; Living arrangement N=597; Relation with the patient N=585.

[d] Education level N=179; Relation with the patient N=175.

[e] Education level N=819; Marital status N=838; Living arrangement N=838; Relation with the patient N=824.

[f] ZBI-12: Zarit Caregiver Burden Interview.

[g] EQ5D: EuroQoL-5D-3L.

The people with dementia/mild cognitive impairment had a mean TechPH index of 2.84 (SD 0.69). The TechEnthusiasm score was 2.95 (SD 1.07) and the TechAnxiety score was 3.30 (SD 0.95). The caregivers had a TechPH index score of 3.07 (SD 0.68); the TechEnthusiasm score was 3.31 (SD 1.00) and the TechAnxiety score was 3.19 (SD 0.95). There were differences in the TechPH index between patients and caregivers (B=0.223, SE 0.030, *P*<.001), which remained significant after adjusting by age, gender, education level, and health status (B=0.074, SE 0.032, *P*=.02). There were significant differences in TechAnxiety between patients and caregivers (B=–0.120, SE 0.041, *P*=0.004); however, these differences were not significant after adjusting for age, gender, education level, and health status (B=–0.042, SE 0.047, *P*=.37). There was a significant difference in TechEnthusiasm between patients and caregivers (B=0.360, SE 0.045, *P*<.001), and this difference was still significant after adjusting for confounders (B=0.128, SE 0.050, *P*=.01). There were significant differences in the use of specifics apps or software to support memory between patients and caregivers (odds ratio [OR]=1.783, SE 0.131, *P*<.001); however, these differences were not significant after adjusting for confounders (OR=0.818, SE 0.167, *P*=.23).

In the multivariate analysis, the TechPH index in people with dementia/mild cognitive impairment was related to lower age, male gender, higher education level, living arrangement (living with children vs single), country of residence (Sweden vs Spain), depression, and health status (Table 3). In caregivers, the TechPH index was associated with lower age, male gender, higher education level, country of residence (Sweden vs Spain), lower caregiver burden, and better quality of life (Table 4). The use of specific apps or software on tablets or mobile phones to support memory was only associated with age in people with dementia/mild cognitive impairment (Table 5) and was associated with age, education level, and caregiver burden in caregivers (Table 6).

Table 3. Multivariate linear regression model[a] of factors associated with technophilia in people with dementia/mild cognitive impairment.

Variables	Estimate	SE	t_{1008}	*P* value	VIF[b]
Intercept	3.763	0.402	9.359	<.001	
Age	–0.009	0.003	–2.850	.004	1.248
Gender: female (male=reference)	–0.160	0.044	–3.630	<.001	1.211
Education level (elementary school=reference)				.01	1.252
Secondary school	0.021	0.052	0.400	.69	
Higher education	0.167	0.058	2.886	.004	
Living arrangement (children=reference)				.01	1.342
Single	–0.210	0.083	–2.538	.01	
Spouse/partner	–0.149	0.074	–1.999	.05	
Other	–0.137	0.103	–1.328	.18	
Country: Sweden (Spain=reference)	0.256	0.058	4.414	<.001	1.855
Diagnosis of dementia: Yes (no=reference)	0.060	0.054	1.116	.26	1.483
Diagnosis of other medical condition: Yes (no=reference)	0.047	0.050	0.939	.35	1.180
Cognitive Status (MMSE[c])	–0.013	0.010	–1.356	.17	1.438
Depression (GDS-15[d])	–0.046	0.008	–5.625	<.001	1.360
Instrumental activities (IADL[e])	–0.020	0.125	–0.164	.87	1.450
Health State (Thermometer)	0.004	0.001	3.570	<.001	1.271

[a] Breusch-Pagan test *P*=.06; Shapiro-Wilk normality test of the residuals *P*=0.39; adjusted R^2=0.146.

[b] VIF: variance inflation factor.

[c] MMSE: Mini Mental State Examination.

[d] GDS-15: Geriatric Depression Scale.

[e] IADL: Lawton Instrumental Activities of Daily Living.

Table 4. Multivariate linear regression model[a] of factors associated with technophilia in caregivers.

Variables	Estimate	SE	t_{1021}	*P* value	VIF[b]
Intercept	3.824	0.159	24.014	<.001	
Age	–0.016	0.002	–9.560	<.001	1.603
Female gender (male=reference)	–0.191	0.043	–4.470	<.001	1.081
Education level (elementary school=reference)				<.001	1.278
Secondary school	0.084	0.050	1.677	.01	
Higher education	0.201	0.052	3.879	<.001	
Living arrangement (children=reference)				.27	1.379
Single	0.050	0.092	0.537	.59	
Spouse/partner	–0.062	0.074	–0.843	.40	
Other	–0.094	0.083	–1.128	.26	
Country: Sweden (Spain=reference)	0.245	0.048	5.140	<.001	1.375
Caregiver Burden (ZBI-12[c])	–0.005	0.003	–2.055	.04	1.154
Quality of life (EQ5D[d])	0.348	0.097	3.592	<.001	1.165

[a] Breusch-Pagan test *P*=.07; Shapiro-Wilk normality test of the residuals *P*=0.84; adjusted R^2=0.194.

[b] VIF: variance inflation factor.

[c] ZBI-12: Zarit Caregiver Burden Interview.

[d] EQ5D: EuroQoL-5D-3L.

Table 5. Multivariate logistic regression model[a] of factors associated with use of apps or specific software to support memory in people with dementia/mild cognitive impairment.

Variables	Estimate	SE	*z*	*P* value	OR[b]	95% CI	VIF[c]
Intercept	3.189	2.247	1.419	.16			
Age	–0.093	0.017	–5.387	<.001	0.911	0.880-0.942	1.366
Gender: female (male=reference)	–0.119	0.233	–0.512	.61	0.877	0.562-1.402	1.183
Education level (elementary school= reference)				.08			1.362
Secondary school	–0.145	0.292	–0.496	.62	0.865	0.488-1.534	
Higher education	0.548	0.290	1.886	.06	1.729	0.979-3.056	
Living arrangement (children=reference)				.16			1.357
Single	1.207	0.663	1.822	.07	3.345	0.912-12.262	
Spouse/partner	1.090	0.625	1.745	.08	2.975	0.874-10.121	
Other	1.347	0.699	1.926	.05	3.844	0.976-15.139	
Country: Sweden (Spain=reference)	0.319	0.312	1.019	.31	1.375	0.745-2.538	1.990
Diagnosis of dementia: Yes (No=reference)	–0.391	0.330	–1.187	.23	0.676	0.354-1.290	1.431
Diagnosis of other medical condition: Yes (No=reference)	0.168	0.262	0.642	.52	1.183	0.708-1.977	1.132
Cognitive Status (MMSE[d])	–0.001	0.056	–0.022	.98	0.999	0.894-1.115	1.373
Depression (GDS-15[e])	–0.008	0.047	–0.165	.87	0.992	0.906-1.087	1.565
Instrumental activities (IADL[f])	0.165	0.744	0.222	.82	1.180	0.274-5.074	1.360
Health State (Thermometer)	0.002	0.006	0.367	.71	1.002	0.990-1.015	1.411

[a] Hosmer and Lemeshow goodness of fit test *P*=.23.

[b] OR: odds ratio.

[c] VIF: variance inflation factor.

[d] MMSE: Mini Mental State Examination.

[e] GDS-15: Geriatric Depression Scale.

[f] IADL: Lawton Instrumental Activities of Daily Living.

Table 6. Multivariate logistic regression model[a] of factors associated with use of apps or specific software to support memory in caregivers.

Variables	Estimate	SE	*z*	*P* value	OR[b]	95% CI	VIF[c]
Intercept	0.089	0.683	0.131	.90			
Age	–0.038	0.007	-5.441	<.001	0.962	0.949-0.976	1.519
Female gender (male=reference)	0.069	0.199	0.346	.73	1.072	0.725-1.585	1.068
Education level (elementary school=reference)				<.001			1.180
Secondary school	0.828	0.252	3.279	.001	2.289	1.395-3.754	
Higher education	0.960	0.257	3.738	<.001	2.611	1.579-4.319	
Living arrangement (children=reference)				.43			1.375
Single	0.080	0.386	0.208	.83	1.083	0.508-2.311	
Spouse/partner	0.332	0.298	1.113	.27	1.394	0.777-2.502	
Other	–0.052	0.337	–0.153	.88	0.950	0.491-1.838	
Country: Sweden (Spain=reference)	–0.126	0.225	–0.561	.57	0.881	0.567-1.370	1.327
Caregiver Burden (ZBI-12[d])	0.023	0.011	2.084	.04	1.023	1.001-1.046	1.141
Quality of life (EQ5D[e])	–0.540	0.429	–1.259	.21	0.582	0.251-1.351	1.194

[a] Hosmer and Lemeshow goodness of fit test *P*=.68.

[b] OR: odds ratio.

[c] VIF: variance inflation factor.

[d] ZBI-12: Zarit Caregiver Burden Interview.

[e] EQ5D: EuroQoL-5D-3L.

Discussion

In people with dementia/mild cognitive impairment, technophilia was associated with less depression, better health status, as well as with sociodemographic variables. In caregivers, technophilia was associated with a better quality of life, less care burden, and other sociodemographic variables. The results of this study indicate that people with dementia/mild cognitive impairment have less technophilia than their caregivers, specifically less TechEnthusiasm, but no differences were found in TechAnxiety. The people with dementia/mild cognitive impairment used smartphones and tablets with specific apps or software to support their memory less than their caregivers, despite being a population who would benefit most from these apps. These differences are likely due to age barriers.

The use of smartphones and tablets in our study was lower than that reported in a recent study in Australia in which 91.4% of people with dementia/mild cognitive impairment reported routinely using smartphones [29]. Likewise, in other studies that included older adults and populations with cognitive impairment and dementia, attitudes toward technology were associated with factors such as age [30-32], male gender [33], higher education level [29,30,33,34], depression and negative cognitions [35], and health status [30], and variability was found between countries [36]. A correlation between the use of health apps and age has also been reported [32,37]. However, some results in other studies have not been replicated. For example, one study reported an association between cognitive functioning and the use of technologies in a cognitively impaired population [38], and another study found that women used more health apps than men [39].

The differences in technophilia between people with dementia/mild cognitive impairment and caregivers, mainly due to the lack of technology enthusiasm, could be related to dementia or mild cognitive impairment itself. In fact, other studies have found that dementia and mild cognitive impairment were related to more perceived difficulties in everyday technology use [40]. In addition, this lack of enthusiasm for technology could be related to the apathy associated with dementia and mild cognitive impairment, which is a persistent behavioral symptom [41,42]. This apathy could affect the enthusiasm toward new technologies and may be a barrier for interventions that demand learning and adaptation to the use of these technologies.

Other studies have also found that barriers to the use of ICTs and assistive technologies in older adults are due to the lack of "interest or relevance to life" and the perception of "no need" [15,43]. Indeed, these ICTs are often not designed specifically to cover the real needs of this population and they do not arouse interest. Other studies have also found that a lack of knowledge in older adults is a barrier to the use of technologies [11,16]. In addition, age was identified as one of the most critical determinants of the use of apps to support memory after adjusting for health status. This result coincides with other studies [15,30] in the general population.

Interestingly, technophilia was also strongly associated with health status and depression after adjusting for possible confounders such as gender, age, education level, and cognitive function. One longitudinal study conducted by Cotten et al [44] and other studies [45,46] have found that use of the internet and ICTs in older individuals is associated with less depression and a better quality of life, and there is some evidence that this association is mediated by loneliness and social isolation [44,45]. In conclusion, these results highlight the potential barriers of interventions for the use of ICTs, including smartphones and tablets, to improve the quality of life of people with dementia/mild cognitive impairment. The nature of the relation between use and technophilia and quality of life and mood needs to be clarified with longitudinal and experimental studies.

The use of specific apps or software to support memory was not associated with better cognitive function, better quality of life, or less depression. However, some positive results have emerged from several meta-analyses reporting that computerized cognitive training and electronic health apps are effective in improving cognition and quality of life in people with dementia/mild cognitive impairment [47,48]. This result comes from formal interventions, and the results of the present study could be due to possible confounders not assessed in the study or because the participants were not using the apps or software adequately (eg, insufficient training time or not using well-designed apps). Other studies have also indicated inconsistencies and lack evidence regarding the effectiveness of ICT interventions to improve cognition and other health-related variables [49,50].

The results of this study need to be considered in the context of several limitations. This was a cross-sectional study, and therefore the causal relation of the variables could not be established. More longitudinal and interventional studies are necessary to determine the effects of ICT use and the attitudes toward technology in people with dementia/mild cognitive impairment and their caregivers and to identify the factors that influence these variables in the other direction. In addition, this was a secondary analysis from a clinical trial, and the sample may not be representative of all people with dementia/mild cognitive impairment and caregivers. Individuals with severe depression were excluded and more than half of the sample was from only one country. There was no established cut-off point for the definition of high or low technophilia, which limits the use of the mean as the cutoff. Another limitation is that only a small set of variables was assessed. However, the strength of the study is the large sample used with more than 2000 individuals.

In conclusion, the factors associated with technophilia suggest potential barriers to technological interventions in people with dementia/mild cognitive impairment and their caregivers. These results have implications on the possible usefulness of considering technophilia and enthusiasm toward ICTs as determinants and moderator elements in digital interventions in the elderly population. The results suggest that designers of apps for older adults with cognitive impairment should create apps that engage users and are designed specifically for their needs as this population frequently lacks enthusiasm for technology. The use of touchscreen technologies was also frequent among people with dementia/mild cognitive impairment, whereas the use of specific apps or software to support memory was reduced in people with dementia/mild cognitive impairment and caregivers, and they face age barriers for the use of these apps. However, the majority of this population considers that use of smartphones or tablets is helpful for memory, which highlights the presence of a gap between the perceived potential and actual use of these technologies.

Acknowledgments

The study is supported by a grant from European Commission, Horizon 2020 (code 643399). This work is a collaboration among several members of the SMART4MD consortium. The consortium acknowledges the contributions of members not reaching full ICMJE standards for authorship in this manuscript: Karen Hayden, Tony Fenton, Jufen Zhang, Selim Cellek, Claire Preston, Tomer Regev, and Stephen Moore (Anglia Ruskin University, UK); Dominique Manhaeve and Jessie Schrijvers (University Colleges Leuven-Limburg, Belgium); Professor María Teresa Arredondo Waldmeyer (Universidad Politécnica de Madrid, UPM, Spain); Ifty Ahmed (Healthbit Ltd, UK); Michal Stefan, Karel Huk, and Jan Kubalik (DEX innovation Center, Czech Republic); Lisa Seeman (Athena ICT, Israel); and Carmen Sanmartin, Line Christensson, Eric Picullel, Joakim Frögren, and Doris Bohman (Blekinge Institute of Technology, Sweden).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed data on use of smartphones and tablets by people with dementia/mild cognitive impairment (PwD/MCI).
[DOCX File , 14 KB - jmir_v22i6e17253_app1.docx]

References

1. World Health Organization and Alzheimer's Disease International. 2012. Dementia: a public health priority Internet URL: https://www.who.int/mental_health/publications/dementia_report_2012/en/ [accessed 2019-11-01]
2. Forrester SN, Gallo JJ, Smith GS, Leoutsakos JS. Patterns of Neuropsychiatric Symptoms in Mild Cognitive Impairment and Risk of Dementia. Am J Geriatr Psychiatry 2016 Feb;24(2):117-125 [FREE Full text] [doi: 10.1016/j.jagp.2015.05.007] [Medline: 26209222]

3. Cerejeira J, Lagarto L, Mukaetova-Ladinska EB. Behavioral and psychological symptoms of dementia. Front Neurol 2012;3:73. [doi: 10.3389/fneur.2012.00073] [Medline: 22586419]
4. Szeto JYY, Lewis SJG. Current Treatment Options for Alzheimer's Disease and Parkinson's Disease Dementia. Curr Neuropharmacol 2016;14(4):326-338 [FREE Full text] [doi: 10.2174/1570159x14666151208112754] [Medline: 26644155]
5. Khosravi P, Ghapanchi AH. Investigating the effectiveness of technologies applied to assist seniors: A systematic literature review. Int J Med Inform 2016 Jan;85(1):17-26. [doi: 10.1016/j.ijmedinf.2015.05.014] [Medline: 26216463]
6. Tyack C, Camic PM. Touchscreen interventions and the well-being of people with dementia and caregivers: a systematic review. Int Psychogeriatr 2017 Aug;29(8):1261-1280. [doi: 10.1017/S1041610217000667] [Medline: 28446258]
7. Berner J, Comijs H, Elmståhl S, Welmer A, Sanmartin Berglund J, Anderberg P, et al. Maintaining cognitive function with internet use: a two-country, six-year longitudinal study. Int Psychogeriatr 2019 Aug 9;31(07):929-936. [doi: 10.1017/s1041610219000668]
8. Koo BM, Vizer LM. Examining Mobile Technologies to Support Older Adults With Dementia Through the Lens of Personhood and Human Needs: Scoping Review. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e15122 [FREE Full text] [doi: 10.2196/15122] [Medline: 31710305]
9. Grossman MR, Zak DK, Zelinski EM. Mobile Apps for Caregivers of Older Adults: Quantitative Content Analysis. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e162 [FREE Full text] [doi: 10.2196/mhealth.9345] [Medline: 30061093]
10. Klimova B. Mobile Phone Apps in the Management and Assessment of Mild Cognitive Impairment and/or Mild-to-Moderate Dementia: An Opinion Article on Recent Findings. Front Hum Neurosci 2017;11:461. [doi: 10.3389/fnhum.2017.00461] [Medline: 28970789]
11. Klimova B, Simonova I, Poulova P, Truhlarova Z, Kuca K. Older people and their attitude to the use of information and communication technologies ? A review study with special focus on the Czech Republic (older people and their attitude to ICT). Educ Gerontol 2016;42(5):361-369. [doi: 10.1080/03601277.2015.1122447]
12. Joe J, Demiris G. Older adults and mobile phones for health: a review. J Biomed Inform 2013 Oct;46(5):947-954 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.008] [Medline: 23810858]
13. Hitch D, Swan J, Pattison R, Stefaniak R. Use of touchscreen tablet technology by people with dementia in homes: A scoping review. J Rehabil Assist Technol Eng 2017;4:2055668317733382 [FREE Full text] [doi: 10.1177/2055668317733382] [Medline: 31186940]
14. Zhang MWB, Chan S, Wynne O, Jeong S, Hunter S, Wilson A, et al. Conceptualization of an evidence-based smartphone innovation for caregivers and persons living with dementia. Technol Health Care 2016 Sep 14;24(5):769-773. [doi: 10.3233/THC-161165] [Medline: 27129032]
15. Selwyn N, Gorard S, Furlong J, Madden L. Older adults' use of information and communications technology in everyday life. Age Soc 2003 Sep 02;23(5):561-582. [doi: 10.1017/S0144686X03001302]
16. O'Connor S, Bouamrane M, O'Donnell CA, Mair FS. Barriers to Co-Designing Mobile Technology with Persons with Dementia and Their Carers. Stud Health Technol Inform 2016;225:1028-1029. [Medline: 27332466]
17. Osiceanu M. Psychological Implications of Modern Technologies: “Technofobia” versus “Technophilia”. Procedia Soc Behav Sci 2015 May;180:1137-1144. [doi: 10.1016/j.sbspro.2015.02.229]
18. Anderberg P, Barnestein-Fonseca P, Guzman-Parra J, Garolera M, Quintana M, Mayoral-Cleries F, et al. The Effects of the Digital Platform Support Monitoring and Reminder Technology for Mild Dementia (SMART4MD) for People With Mild Cognitive Impairment and Their Informal Carers: Protocol for a Pilot Randomized Controlled Trial. JMIR Res Protoc 2019 Jun 21;8(6):e13711 [FREE Full text] [doi: 10.2196/13711] [Medline: 31228177]
19. Anderberg P, Eivazzadeh S, Berglund JS. A Novel Instrument for Measuring Older People's Attitudes Toward Technology (TechPH): Development and Validation. J Med Internet Res 2019 May 23;21(5):e13951 [FREE Full text] [doi: 10.2196/13951] [Medline: 31124467]
20. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Quality of life in Alzheimer's disease: Patient and caregiver reports. Journal of Mental Health and Aging 1999;5(1):21-32 [FREE Full text]
21. van Reenen M, Oppe M. EQ-5D. 2015. EQ-5D-3L user guide URL: https://euroqol.org/publications/user-guides/ [accessed 2020-05-10]
22. König H, Bernert S, Angermeyer MC, Matschinger H, Martinez M, Vilagut G, ESEMeD/MHEDEA 2000 Investigators. Comparison of population health status in six european countries: results of a representative survey using the EQ-5D questionnaire. Med Care 2009 Feb;47(2):255-261. [doi: 10.1097/MLR.0b013e318184759e] [Medline: 19169128]
23. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]
24. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
25. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
26. World Health Organization. 2004. ICD-10 : international statistical classification of diseases and related health problems : tenth revision, 2nd ed URL: https://apps.who.int/iris/handle/10665/42980 [accessed 2020-05-10]

27. Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. Gerontologist 1980 Dec;20(6):649-655. [doi: 10.1093/geront/20.6.649] [Medline: 7203086]
28. Bédard M, Molloy DW, Squire L, Dubois S, Lever JA, O'Donnell M. The Zarit Burden Interview: a new short version and screening version. Gerontologist 2001 Oct;41(5):652-657. [doi: 10.1093/geront/41.5.652] [Medline: 11574710]
29. LaMonica HM, English A, Hickie IB, Ip J, Ireland C, West S, et al. Examining Internet and eHealth Practices and Preferences: Survey Study of Australian Older Adults With Subjective Memory Complaints, Mild Cognitive Impairment, or Dementia. J Med Internet Res 2017 Oct 25;19(10):e358 [FREE Full text] [doi: 10.2196/jmir.7981] [Medline: 29070481]
30. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
31. Halmdienst N, Radhuber M, Winter-Ebmer R. Attitudes of elderly Austrians towards new technologies: communication and entertainment versus health and support use. Eur J Ageing 2019 Dec;16(4):513-523 [FREE Full text] [doi: 10.1007/s10433-019-00508-y] [Medline: 31798375]
32. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
33. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421. [doi: 10.1093/geront/gnt166] [Medline: 24379019]
34. Tacken M, Marcellini F, Mollenkopf H, Ruoppila I, Széman Z. Use and acceptance of new technology by older people. Findings of the international MOBILATE survey: 'Enhancing mobility in later life.'. Gerontechnology 2005;3(3):126-137. [doi: 10.4017/gt.2005.03.03.002.00]
35. Conrad AM, Munro D. Relationships between Computer Self-Efficacy, Technology, Attitudes and Anxiety: Development of the Computer Technology Use Scale (CTUS). J Educ Comput Res 2008 Oct 20;39(1):51-73. [doi: 10.2190/ec.39.1.d]
36. König R, Seifert A, Doh M. Internet use among older Europeans: an analysis based on SHARE data. Univ Access Inf Soc 2018 Jan 19;17(3):621-633. [doi: 10.1007/s10209-018-0609-5]
37. Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, et al. Prevalence of Health App Use Among Older Adults in Germany: National Survey. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e26 [FREE Full text] [doi: 10.2196/mhealth.8619] [Medline: 29362211]
38. Hedman A, Nygård L, Almkvist O, Kottorp A. Amount and type of everyday technology use over time in older adults with cognitive impairment. Scand J Occup Ther 2015 May;22(3):196-206. [doi: 10.3109/11038128.2014.982172] [Medline: 25580938]
39. Xie Z, Nacioglu A, Or C. Prevalence, Demographic Correlates, and Perceived Impacts of Mobile Health App Use Amongst Chinese Adults: Cross-Sectional Survey Study. JMIR Mhealth Uhealth 2018 Dec 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
40. Rosenberg L, Kottorp A, Winblad B, Nygård L. Perceived difficulty in everyday technology use among older adults with or without cognitive deficits. Scand J Occup Ther 2009;16(4):216-226. [doi: 10.3109/11038120802684299] [Medline: 19148847]
41. Sherman C, Liu CS, Herrmann N, Lanctôt KL. Prevalence, neurobiology, and treatments for apathy in prodromal dementia. Int Psychogeriatr 2018 Feb;30(2):177-184. [doi: 10.1017/S1041610217000527] [Medline: 28416030]
42. van der Linde RM, Matthews FE, Dening T, Brayne C. Patterns and persistence of behavioural and psychological symptoms in those with cognitive impairment: the importance of apathy. Int J Geriatr Psychiatry 2017 Mar;32(3):306-315. [doi: 10.1002/gps.4464] [Medline: 27017917]
43. Yusif S, Soar J, Hafeez-Baig A. Older people, assistive technologies, and the barriers to adoption: A systematic review. Int J Med Inform 2016 Oct;94:112-116. [doi: 10.1016/j.ijmedinf.2016.07.004] [Medline: 27573318]
44. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):763-771. [doi: 10.1093/geronb/gbu018] [Medline: 24671896]
45. Khalaila R, Vitman-Schorr A. Internet use, social networks, loneliness, and quality of life among adults aged 50 and older: mediating and moderating effects. Qual Life Res 2018 Feb;27(2):479-489. [doi: 10.1007/s11136-017-1749-4] [Medline: 29210015]
46. Lee HY, Kim J, Sharratt M. Technology use and its association with health and depressive symptoms in older cancer survivors. Qual Life Res 2018 Feb;27(2):467-477. [doi: 10.1007/s11136-017-1734-y] [Medline: 29128998]
47. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
48. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
49. Butler M, McCreedy E, Nelson VA, Desai P, Ratner E, Fink HA, et al. Does Cognitive Training Prevent Cognitive Decline?: A Systematic Review. Ann Intern Med 2018 Jan 02;168(1):63-68. [doi: 10.7326/M17-1531] [Medline: 29255842]

50. Ge S, Zhu Z, Wu B, McConnell ES. Technology-based cognitive training and rehabilitation interventions for individuals with mild cognitive impairment: a systematic review. BMC Geriatr 2018 Sep 15;18(1):213 [FREE Full text] [doi: 10.1186/s12877-018-0893-1] [Medline: 30219036]

Abbreviations

EQ5D: EuroQoL-5D-3L
EQ-VAS: Scale thermometer of 0-100
GDS-15: Geriatric Depression Scale
IADL: Lawton Instrumental Activities of Daily Living
ICT: information and communication technology
MMSE: Mini-Mental State Examination
OR: odds ratio
QoL-AD: Quality of Life in Alzheimer's Disease
SMART4MD: Support Monitoring and Reminder for Mild Dementia
VIF: variance inflation factor
ZBI-12: Zarit Caregiver Burden Interview

Edited by G Eysenbach; submitted 28.11.19; peer-reviewed by S Yusif, J Synnott, A Olsson; comments to author 09.01.20; revised version received 18.02.20; accepted 22.03.20; published 01.06.20.

Please cite as:
Guzman-Parra J, Barnestein-Fonseca P, Guerrero-Pertiñez G, Anderberg P, Jimenez-Fernandez L, Valero-Moreno E, Goodman-Casanova JM, Cuesta-Vargas A, Garolera M, Quintana M, García-Betances RI, Lemmens E, Sanmartin Berglund J, Mayoral-Cleries F
Attitudes and Use of Information and Communication Technologies in Older Adults With Mild Cognitive Impairment or Early Stages of Dementia and Their Caregivers: Cross-Sectional Study
J Med Internet Res 2020;22(6):e17253
URL: https://www.jmir.org/2020/6/e17253
doi:10.2196/17253
PMID:32442136

Original Paper

Gender and the Digital Divide Across Urban Slums of New Delhi, India: Cross-Sectional Study

Ashish Joshi[1], PhD, MBBS, MPH; Bhavya Malhotra[2], MSc; Chioma Amadi[1], MPH; Menka Loomba[2], MPH; Archa Misra[2], MSc; Shruti Sharma[2], MSc; Arushi Arora[2], MPH; Jaya Amatya[1], MPH

[1]Graduate School of Public Health and Health Policy, City University of New York, New York, NY, United States
[2]Foundation of Health care Technologies Society, Delhi, India

Corresponding Author:
Ashish Joshi, PhD, MBBS, MPH
Graduate School of Public Health and Health Policy
City University of New York
55W 125th Street
New York, NY
United States
Phone: 1 4435706018
Email: ashish.joshi@sph.cuny.edu

Abstract

Background: Disparities in access to specific technologies within gender groups have not been investigated. Slum settings provide an ideal population to investigate the contributing factors to these disparities.

Objective: This study aimed to examine gender differences in mobile phone ownership, internet access, and knowledge of SMS text messaging among males and females living in urban slum settings.

Methods: A convenience sampling approach was used in sample selection from 675 unnotified slums. A total of 38 slum sites were then selected across four geographic zones. Of these, 10% of the households in each slum site was selected from each zone. One household member was interviewed based on their availability and fulfillment of the eligibility criteria. Eligible individuals included those aged 18 years and above, residing in these slums, and who provided voluntary consent to participate in the study. Individuals with mental or physical challenges were excluded from the study.

Results: Our results showed that females were half as likely to own mobile phones compared with males (odds ratio [OR] 0.53, 95% CI 0.37-0.76), less likely to have internet access (OR 0.79, 95% CI 0.56-1.11), or know how to send text messages (OR 0.93, 95% CI 0.66-1.31). The predictors of mobile phone ownership, internet access, and text messaging *between* males and females included age, individual education, housing type, and the number of earning members in a household in the adjusted analysis. Among males, the number of earning members was a predictor of both mobile phone ownership and text messaging, whereas household education was a predictor of both internet access and text messaging. Age and individual education only predicted internet access, whereas housing type only predicted text messaging. Among females, household education was a predictor of all the technology outcomes. Age and type of toilet facility only predicted mobile phone ownership; housing type only predicted internet access whereas television ownership with satellite service and smoking behavior only predicted text messaging.

Conclusions: Our study findings showing disparate access to technology *within* gender groups lend support for further research to examine the causal mechanisms promoting these differences to proffer significant solutions. Specifically, our study findings suggest that improving household education is crucial to address the disparate access and usage of mobile phones, the internet, and text messaging among women in slum settings. This suggestion is due to the consistency in household educational level as a predictor across all these technology indicators. In addition, the mechanisms by which the number of household earning members influences the disparate access to technology among men call for further exploration.

(J Med Internet Res 2020;22(6):e14714) doi:10.2196/14714

KEYWORDS

gender; digital divide; mobile phone; internet access; text messaging; slums

Introduction

Digital Divide

Information and Communication Technology (ICT) plays a major role in fostering access to knowledge and key services across various sectors. ICT proliferation has been associated with increasing economic benefits, including new economic opportunities, increased trade, higher productivity, and lower costs. ICT is in a continuous state of advancement, evolution, and rapid diffusion at record-setting rates. For instance, global internet penetration rose from around 6% to almost 50% between 2000 and 2016, with penetration rates exceeding 90% in developed countries [1]. The term *digital divide* was coined in the 1990s by Lloyd Morriset (President of the Markle Foundation) to describe these inequalities, which depict a divide between the *information-haves and have-nots* [2]. The digital divide connotes disparate access to information across individuals with and without access to the internet, and more broadly, ICT as well as the general media [2]. However, the digital divide is most commonly used for indicating the availability of internet access at an affordable cost and quality [3,4]. Disparate patterns in technology access are frequently measured by internet access, penetration, the number of internet users, household ownership of computers, and mobile phone usage [5].

Background

The factors impacting the digital divide were conceptualized using the resources and appropriation theory developed by Van Dijk [5]. This theoretical framework depicts a causal model of the interplay between 1) individual and societal inequalities, 2) distribution of resources, 3) access to ICTs, and 4) societal participation. Specifically, the differential growth in ICT access and usage across countries in varied settings has been attributed to gross inequalities at the individual and societal levels. These are observed across age distributions, gender, race or ethnicity, income, literacy, personality, health, household conditions, and socioeconomic status (SES) [5]. The inequalities subsequently produce an unequal distribution of resources that promote unequal access to technology. Unequal access to technology is also dependent on the type of technology (ie, basic phone with limited functionality such as the absence of cameras, smartphone, computer, and other advanced systems). Unequal access to technology contributes to unequal societal participation, which, in turn, reinforces the existing inequalities [5]. Such inequalities tend to be more pronounced in marginalized settings, notably in urban slum settings in developing countries, which constitute a hub of economic disadvantage.

Although a variety of individual-level determinants of ICT inequalities exist, gender remains one that is of primary importance, as reflected in the United Nations (UN) Sustainable Development Goal 5: Gender Equality. The Sustainable Development Goal 5B target is to enhance the use of enabling technology, in particular ICT technology, to promote the empowerment of women [6]. Unequal access to technology among men and women constitutes one of the most striking aspects of the digital divide. The impact of the digital divide on gender has been widely studied in various developing and developed economies [7]. According to the International Telecommunication Union (ITU), on average, women are 16 percentage points less likely to use the internet compared with their male counterparts. This gender-gap is consistent globally, varying between 11% and 19% in Nigeria, Tanzania, India, Pakistan, and Japan, with differences as high as 31% in the least developed settings [7]. Gender gaps in these settings have been attributed to a variety of determinants, including disproportionate access to education among young women, which impairs their literacy levels. Further determinants include limited institutional opportunities for ICTs, personal safety issues with access to ICTs, and the *leaky pipe phenomenon*, which describes female preference to advance their family’s welfare over their personal development [6]. These barriers are perpetuated by structural factors such as extreme poverty and highly patriarchal societies, as well as psychological barriers such as limited confidence among women in their capacity to learn ICT skills, and the belief that technology is reserved for their male counterparts [7].

Objectives

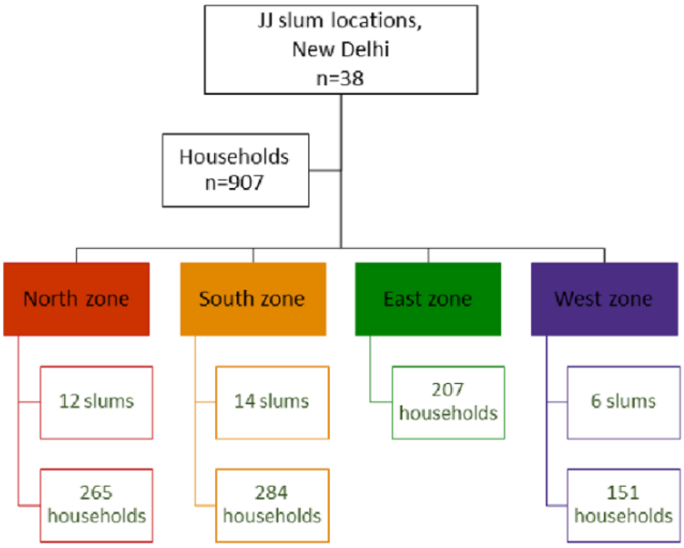
Although prior research has established the existence of a digital divide between gender categories, findings of these studies suggest that variations within individual gender categories may be significantly contributing to the existing digital divide [8]. Slum settings provide a unique population to investigate the contributing factors to the disparity in technology access and usage within gender categories. This is because slums represent a hub of staggering economic disparities, which tend to be more diverse than the nonslum populations. In particular, a UN habitat report indicated that slums suffer from higher disease incidence and mortality, which exceed nonslum populations, and these disparities are rarely reflected in the national statistics, thereby masking the extent of the deprivation in slum settings [9]. Essentially, the state of slums constitutes an indicator of prosperous cities [9]. The objective of this study was to examine gender differences in mobile phone ownership, internet access, and knowledge of SMS among males and females living in urban slum settings.

Methods

Overview

A cross-sectional study was conducted between June 2016 and January 2017 to assess the impact of the digital divide among women and men residing in urban slums of Delhi, India. The sampling frame used was the *Delhi Urban Shelter Improvement Board Jhuggi-Jhopadi Cluster List of 2015*, which enumerated a total of 675 unnotified urban slums across the four geographic zones (North, South, East, and West) of New Delhi, India. Unnotified slums are slum settlements that are not federally recognized and do not benefit from government subsidies and interventions across slums [10]. A convenience sampling approach was used in identifying 675 unnotified slums, and selecting 38 slum sites across four zones (north zone, n=12; south zone, n=14; east zone, n=6; and west zone, n=6; Figure 1). From each zone, 10% of the households were selected based on proximity to the researcher, ease of access to the slums, and the presence of local collaborators in these slums who could help in navigating them. Of these, 1 member from each household was selected and interviewed based on availability for the interview and fulfillment of the eligibility criteria. Eligible individuals included those who were aged 18 years and above, resident in these slums, and provided voluntary consent to participate in the study. Individuals who did not provide consent and had any mental challenges were excluded from the study. This resulted in a total sample of 907 study participants across all the slums. The Institutional Review Board (IRB) of the Foundation of Healthcare Technologies Society, New Delhi, India, approved the study protocol (IRB number: FHTS/041/2016).

Figure 1. Study participant recruitment.



Variables Assessed

The following variables were assessed:

- Sociodemographic characteristics: age, gender, education, household education, type of family, earning members in the household, housing type, type of toilet facility, television ownership, and healthy behaviors, including smoking and alcohol consumption.
- Living index: information was collected about housing type (concrete, semiconcrete, or not concrete), access to toilet facility (in-house, public place, or open defecation), television ownership, and use of satellite television service.
- High-risk behaviors: information was collected on the reporting of high-risk behaviors, including smoking and alcohol consumption.
- Mobile ownership, internet access, and knowledge of SMS: information was collected about individual or household mobile phone ownership, internet service (mobile phone), and knowledge of SMS text messaging.

Statistical Analysis

Descriptive analysis was conducted to report means with SDs and frequencies for all continuous and categorical variables, respectively. Association was performed between sociodemographic characteristics, living index, and health behaviors and technology outcomes, including mobile phone ownership, access to the internet, and knowledge of text messaging. Stratified analysis by gender was performed to determine the between sociodemographic characteristics, living index, and health behaviors and technology outcomes, including mobile phone ownership, access to the internet, and knowledge of text messaging. Variables having significant relationship were included in multivariable logistic regression. Multivariate analyses, stratified by gender, were performed to examine variables that were associated with mobile phone ownership, internet access, and knowledge of SMS after adjusting for potential confounders including sociodemographics, healthy behaviors, and living index conditions. The analyses were performed using SAS, AS, Version 9.4 (SAS Software Limited).

Results

Study Participant Characteristics

The average age of the study participants was 36 years (SD 13). Almost half of them were between the ages of 18 and 30 years (398/904, 44.0%). More than half of them were females (599/904, 66.3%), and 46.2% (418/904) had not completed high school. Half of them lived in households where the highest level of education was less than a high school diploma (453/904, 50.1%). More than half of the study participants lived in nuclear families (578/904, 64%) and had 1 earning member per household (534/904, 59.1%).

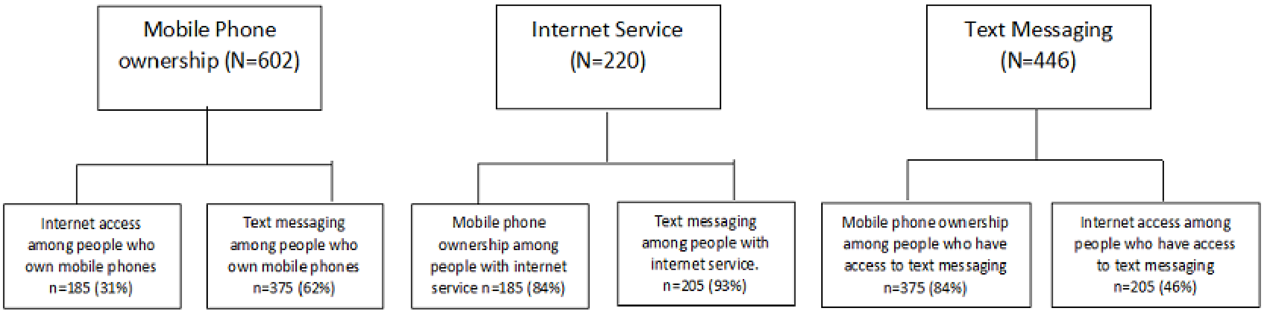
More than half of the study participants resided in houses that had concrete finishing (496/904, 54.9%). Of which, 45.2% (409/904) of the study participants utilized public toilet facilities as their primary source of sanitation. More than two-thirds of the study participants owned a television set (705/904, 77.9%) and had satellite television service (592/904, 65.5%). Then, 22.0% (199/904) of the participants reported smoking and 11.6% (105/904) reported alcohol consumption. The characteristics of the study participants have been previously published [11]. Less than 1% of responses were missing across the technology outcomes assessed (34/85000, 0.04%).

Mobile Phone Ownership, Internet Service, and Text Messaging Among Study Participants

More than half of the study participants owned a mobile phone (602/904, 66.5%). Only 24.3% (220/904) of the study participants had internet service on their mobile phones. Less than half of the study participants had knowledge of sending text messages (446/904, 49.3%; Figure 2).

More than half of the study participants who owned mobile phones had knowledge of sending text messages (375/602, 62.2%). However, less than half of them had internet service on their mobile phones (185/602, 30.7%; Figure 2 and Multimedia Appendix 1). More than two-thirds of the study participants who had internet service owned a mobile phone (185/220, 84.1%), and 93.2% (205/220) of them had knowledge of sending text messages. In all, 84.1% (375/446) of the study participants who knew how to send text messages owned a mobile phone, and less than half of them had access to the internet (205/446, 45.9%; Figure 2).

Figure 2. Phone ownership, internet service, and text messaging among the study participants (N=904).



Mobile Phone Ownership and Technology Access and Familiarity

Variables Associated With Mobile Phone Ownership, Internet Access, and Knowledge of Text Messaging Among the Study Participants

Mobile Phone Ownership

More than half of the study participants owned a mobile phone (602/904, 66.3%). A total of 61.7% (365/904) of them were females, 43.8% (264/602) of them were between 18 and 30 years old, and 54.3% (327/602) had not obtained a high school diploma (Multimedia Appendix 1). More than half of them lived in nuclear families having 1 earning member (351/602, 58.3%), and with a household educational level of less than high school (312/602, 51.8%). College education at the individual and household levels were significantly higher among study participants who owned a mobile phone compared with those that did not (*P*<.001).

A total of 61.7% (372/602) of the study participants that owned a mobile phone resided in households with concrete finishing compared with 41.1% (124/302) of those who did not own a mobile phone (*P*<.001). Less than 10% of the participants that owned a mobile phone practiced open defecation (40/602, 6.6%) compared with 25.8% (78/302) of those that did not own a mobile phone (*P*<.001). The majority of mobile phone owners also owned a television (515/602, 85.5%) and had access to a satellite television service (437/602, 72.5%). Smoking (*P*=.50) and alcohol consumption (*P*=.40) were not significantly associated with those owning a mobile phone (Multimedia Appendix 1). Age of the study participant, gender, individual and household education, type of family, number of earning members, housing type, type of toilet facility, and television ownership with satellite service were significantly associated with mobile phone ownership.

Internet Access

Less than one-third of the study participants had internet service on their mobile phones or had a household member with internet service (220/904, 24.3%). percent of these, 51.4% (113/220) of those with internet service on their phones were between the ages of 18 and 30 years (*P*=.02; Multimedia Appendix 1). More than half of them were females (127/220, 57.7%) and had not completed high school (116/220, 52.7%). In all, 52.7% (116/220) of them lived in nuclear family settings and had 1 earning family member (110/220, 50.0%). More than two-thirds of them resided in houses with concrete finishing (153/220, 69.5%), and 50.0% (111/220) of them utilized in-house toilet facilities. Then, 78.6% (173/220) of them owned a television with satellite television service (*P*<.001)

The study participants age (*P*=.02), gender (*P*=.002), individual and household education, type of family, number of earning members, housing type (*P*=.02), type of toilet facility, and television ownership with satellite service were significantly associated with internet access on a mobile phone (*P*<.001). Smoking (*P*=.50) and alcohol consumption (*P*=.40) were not significantly associated with having internet service on their mobile phones.

Text Messaging

Less than half of the study participants knew how to send text messages (446/904, 49.3%; Multimedia Appendix 1). More than half of them were females (276/446, 61.9%) and had not completed high school (252/446, 56.5%). Almost one-third of the study participants who were knowledgeable about texting had obtained some college education (123/446, 27.5%) compared with 7.0% (31/446) of those who did not know how to send text messages. The majority of them had 1 to 2 earning members in the household. More than two-thirds of them resided in houses with concrete finishing (299/446, 67.0%). Less than half of them utilized public places (205/446, 45.9%) or in-house toilet facilities (212/446, 47.5%) as their primary mode of sanitation. Of these, 6.5% (29/446) practiced open defecation. The majority of them owned a television (402/446, 90.1%) and had satellite television service (347/446, 77.8%). Smoking and alcohol consumption were not significantly associated with the knowledge of text messaging. Age of the study participant, gender, individual and household education, type of family, number of earning members, housing type, type of toilet facility, television ownership with satellite service, and technology usage were significantly associated with the knowledge of text messaging.

Predictors of Mobile Phone Ownership, Internet Access, and Text Messaging Among the Study Participants

Predictors of Mobile Phone Ownership

Age, household educational level, number of earning members, housing type, and type of toilet facility remained significantly associated with mobile phone ownership (Multimedia Appendix 2). Females had lower odds of mobile phone ownership compared with males (odds ratio [OR] 0.51, 95% CI 0.35-0.75). Study participants in the age group of 41-50 years had the highest odds of mobile phone ownership as compared with those in the 50+ age (OR 1.99, 95% CI 1.07-3.69). Individuals living in households where the highest educational level was less than a high school diploma was associated with lower odds of mobile phone ownership (OR 0.51, 95% CI 0.28-0.94). Similarly, no earning member in a household was associated with lower odds of mobile phone ownership (OR 0.27, 95% CI 0.08-0.91). Living in houses made of concrete was associated with higher odds of mobile phone ownership (OR 1.99, 95% CI 1.13-3.49). In addition, living in households that utilized public places as their primary mode of sanitation was associated with a higher odds of mobile phone ownership (OR 1.56, 95% CI 1.07-2.25), whereas open defecation was associated with a lower odds of mobile phone ownership (OR 0.49, 95% CI 0.29-0.84).

Predictors of Internet Access

Age and household education remained significantly associated with internet access (Multimedia Appendix 2). Females had lower odds of internet access compared with males (OR 0.65, 95% CI 0.44-0.96). Study participants aged between 18 and 30 years had the highest odds of internet access across age groups (OR 2.19, 95% CI 1.15-4.19). Households in which the highest level of education attained was high school were associated with a lower odds of internet access (OR 0.49, 95% CI 0.28-0.87).

Predictors of Text Messaging

Age, individual education, household education, number of earning members in the household, housing type, and satellite television service remained significantly associated with text messaging (Multimedia Appendix 2). Study participants between the ages of 41 and 50 had the highest odds of text messaging (OR 2.12, 95% CI 1.13-3.97). Households in which no member had obtained any schooling was associated with lower odds of text messaging (0.22, 95% CI 0.05-0.93). Living in houses made of concrete (OR 2.05, 95% CI 1.13-3.72) and having a satellite television service were associated with higher odds of text messaging (OR 1.94, 95% CI 1.12-3.33).

Stratified Analysis of Mobile Phone Ownership, Internet Access, and Text Messaging Across Gender Categories (Bivariable Analysis)

Mobile Phone Ownership Stratified by Gender

A total of 39.4% (237/602) of males and 60.1% (365/602) of females owned mobile phones. Gender differences in mobile phone ownership were significant across age groups (*P*=.002), educational status of the participant (*P*<.001), housing type (*P*=.02), smoking (*P*=.001), and alcohol consumption (*P*=.02; Multimedia Appendix 3). Of which, 60.1% (365/602) of the females owned mobile phones compared with 39.3% (237/602) males; 8.4% (20/237) of the males who owned a mobile phone had obtained some college education compared with 5.2% (19/365) of the females (*P*<.001); 8.5% (31/365) of the females who owned a mobile phone resided in houses with nonconcrete finishing compared with 5.1% (12/237) of the males (*P*=.02). One-third of the males who owned a mobile phone reported smoking (n=73), compared with 17.3% (63/365) of the females. And 16.9% (40/237) of the males reported alcohol consumption compared with 10.4% (38/365) of the females. Gender differences were not significant by household education (*P*=.74), family type (*P*=.97), total earning members (*P*=.19), type of toilet facility (*P*=.91), and television ownership (*P*=.29)

Access to Internet Services Stratified by Gender

A total of 42% (93/220) of the males had access to internet services as compared with 57.7% (127/220) of females. Gender differences in internet access were significant across educational levels (*P*<.001) and housing type (*P*=.01; Multimedia Appendix 4). Almost one-quarter of the males had obtained some college education (21/93, 23%) compared with 9% (11/127) of females (*P*<.001) More than two-thirds of the females (94/127, 74%) resided in houses made of concrete compared with 64% (59/93) of the males. Gender differences in access to the internet were not significant by age (*P*=.09), household education (*P*=.25), family type (*P*=.78), toilet facility (*P*=.73), television ownership (*P*=.09), smoking (*P*=.05), and alcohol consumption (*P*=.92) .

Text Messaging Stratified by Gender

A total of 38.1% (170/446) of males were familiar with sending text messaging as compared with 61.8% (276/446) of females. Gender differences were significant across educational groups (*P*<.001), total earning members in the household (*P*=.03), housing type (*P*<.001), smoking behaviors (*P*<.001), and alcohol consumption (*P*<.04; Multimedia Appendix 5). Almost half of the study participants capable of texting were aged 18 to 30 years. Of which 47.1% (130/276) of them were females compared with 45.9% (78/170) of them that were males; 14.2% (24/170) of the males were 50 years or older compared with 7.9% (22/276; *P*=.03) of females; and 12.3% (21/170) of the males had obtained some college education compared with 7.6% (21/276) of the females (*P*<.001). Less than half of the males had 1 earning member (83/170, 48.8%) compared with 58.7% (162/276) of the females. More than two-thirds of the females (201/276, 72.8%) resided in houses made of concrete compared with 57.6% (98/170) of the males, 30.6% (52/170) of the males reported smoking (*P*<.001), and 15.9% (27/170) of the males reported drinking (*P*=.04).

Factors Associated With Mobile Phone Ownership, Internet Access, and Text Messaging Stratified by Gender (Multivariable Analysis)

Age (*P*=.002), educational status (*P*=.003), and type of housing (*P*=.004) remained predictors of mobile phone ownership between males and females (Multimedia Appendix 6). Being older than 30 years (OR 2.19, 95% CI 1.35-3.58) and living in houses made of concrete (OR 2.49, 95% CI 1.50-4.14) were also significant predictors of mobile phone ownership. Educational status was the only significant predictor of internet access between males and females (*P*<.001). Older age (*P*=.004), less than high school education (*P*<.001), 1 household earning member, or less (*P*=.02) were associated with a lesser odds of text messaging between males and females. Living in houses made of concrete was associated with greater odds of text messaging (OR 2.69, 95% CI 1.55-4.69) (Multimedia Appendix 6).

Within-Gender Variation Related to Mobile Phone Ownership, Internet Access, and Text Messaging

Within-Gender Variation Related to Mobile Phone Ownership

Educational status of the individual (*P*<.001), household educational attainment (*P*<.001), total earning members in the household (*P*<.001), housing type (*P*<.001), type of toilet facility (*P*<.001), television ownership (*P*<.001), having a satellite television service (*P*<.001), and smoking (*P*=.02) were significantly associated with mobile phone ownership among males (Multimedia Appendix 7). More than half of the males who owned mobile phones had not obtained a high school education (154/237, 64.9%); resided in households where the highest level of education was less than high school (127/237, 53.6%), and had 1 earning member (129/237, 54.4%). More than half of them resided in houses with concrete finishing (137/237, 57.8%), and half of them utilized public places as the main source of sanitation (n=119). More than two-thirds of them owned a television and had satellite television service. One-third of the male participants who owned a mobile phone were smokers (n=73/237). Age (*P*=.30), type of family (*P*=.31), and alcohol consumption (*P*=.89) were not significantly associated with mobile phone ownership among men (Multimedia Appendix 7).

Age of the study participants, individual and household educational attainment, type of family, housing type, type of toilet facility, television ownership, and having a satellite television service were significantly associated with mobile phone ownership among female study participants (*P*<.001; Multimedia Appendix 6). Almost half of the females who owned a mobile phone were between 18 and 30 years old (167/365, 45.2%) and had not obtained a high school diploma (173/365, 47.4%; *P*<.001). More than half of them resided in households where the highest level of education attained was less than a high school diploma (185/365, 50.7%), and 64% (234/365) of them lived in nuclear families and in houses with concrete finishing (n=235). Less than half of them utilized public places as their main source of sanitation (177/365, 48.5%); 84.9% (310/365) of them owned a television, and 70.1% (256/365) of them had satellite television service. The number of earning members in the household (*P*=.52), smoking (*P*=.24), and alcohol consumption (*P*=.14) were not significantly associated with mobile phone ownership among women (Multimedia Appendix 7).

Within-Gender Variation Related to Internet Access

Age of the study participant (*P*=.002), educational level (*P*<.001), household education (*P*<.001), total earning members in the household (*P*=.003), and housing type (*P*=.01) were significantly associated with mobile phone internet access among men (Multimedia Appendix 8). More than half of the males who had internet access on their mobile phones were aged 18 to 30 years (*P*=.002). Almost a quarter of them had obtained some college education (21/93, 23%). Almost half of them lived in households where the highest level of education attained was college (40/93, 43%), and 41% (38/93) of them had 1 earning member in the household.

The educational level of the study participant, household education, type of family, housing type, type of toilet facility, television ownership, and satellite television service were significantly associated with mobile phone internet access among the females (*P*<.05). Of these, 9.0% (11/122) of females who had internet access had obtained some college education (*P*<.001). More than one-third of the females with internet access lived in households where a college degree had been attained (45/127, 35.4%; *P*<.001). Half of them came from nuclear families (64/127, 50.4%). More than two-thirds resided in houses made of concrete (94/127, 74.0%) and primarily used in-house toilet facilities (67/127, 52.8%). Almost all of them owned a television (117/127, 92.1%) and had satellite television service (103/127, 81.1%). Age (*P*=.22), total earning members in the household (*P*=.33), smoking (*P*=.29), and alcohol consumption (*P*=.05) were not significantly associated with internet access among females (Multimedia Appendix 8).

Within-Gender Variation Related to Text Messaging

The study participant's education (*P*<.001), household education (*P*<.001), total earning members (*P*=.001), housing type (*P*<.001), toilet facility (*P*<.001), television ownership (*P*<.001), and satellite television service (*P*<.001) were significantly associated with the knowledge of text messaging among males (Multimedia Appendix 9). More than half of the males who were familiar with text messages had not completed high school (112/170, 65.9%; *P*<.001). Half of them lived in households where the highest level of education attained was less than a high school diploma (85/170, 50.0%; *P*<.001). Less than half of them had 1 earning member (83/170, 48.8%), 57.6% (98/170) of them lived in houses made of concrete, and 50.0% (85/170) of them utilized public places as their primary source of sanitation. The majority of them owned a television (153/170, 90.0%) and had satellite television service (134/170, 78.8%; *P*<.001). High-risk behaviors, including smoking (*P*=.15) and alcohol consumption (*P*=.66) were not significantly associated with text messaging among males.

Age of the study participant (*P*=.02), education (*P*<.001), household education (*P*<.001), type of family (*P*=.002), housing type (*P*<.001), toilet facility (*P*<.001), television ownership

(*P*<.001), satellite television service (*P*<.001), and smoking (*P*=.03) were significantly associated with the knowledge of text messaging among females (Multimedia Appendix 9). Almost half of the females who were familiar with sending text messages were between the ages of 18 and 30 years (130/276, 47.1%; *P*=.02). More than half of them had not completed high school (140/276, 50.7%), and less than half of them resided in households where the highest level of education was less than a high school diploma (121/276, 43.8%; *P*<.001). More than half of them resided in nuclear families (158/276, 57.2%) with 1 earning member (162/276, 59%). Half of them lived in houses with in-house toilet facilities; 72.8% (201/276) of the females who were familiar with sending text messaging resided in houses made of concrete compared with 43.0% (139/323) who were not capable of texting (*P*<.001). The majority of them owned a television (249/276, 90.2%) and had satellite television service (213/276, 77.1%); 12.3% (34/276) of females who were capable of texting reported smoking behaviors compared with 18.9% (61/323) who were not capable of texting (*P*=.03).

Multivariable Analysis of Factors Associated With Mobile Phone Ownership Within Males and Females

The number of household-earning members and the type of toilet facility in a household remained significantly associated with mobile phone ownership among males (*P*=.01; Table 1). Males who had no earning members in the household had a lesser odds of mobile phone ownership (OR 0.06, 95% CI 0.01-0.54; *P*=.01), and a higher odds of practicing open defecation as their primary source of sanitation (OR 2.88, 95% CI 1.09-7.63; *P*=.03).

Age of the study participant (*P*=.01), household education (*P*=.001), and type of toilet facility (*P*=.03) remained significantly associated with mobile phone ownership among females. Females between the ages of 31 and 40 years had the highest odds of mobile phone ownership among all the age groups (OR 2.418, 95% CI 1.243-4.703; *P*=.009). Females in households where no one had obtained any schooling had the lowest odds of mobile phone ownership (OR 0.27, 95% CI 0.12-0.59). In addition, females living in households that utilized public places as their primary mode of sanitation had a higher odds of mobile phone ownership (OR 1.59, 95% CI 1.04-2.43; *P*=.03), whereas those practicing open defecation had a lower odds of mobile phone ownership (OR 0.47, 95% CI 0.26-0.87).

Table 1. Multivariable analysis showing predictors of mobile phone ownership within the male and female study participants (N=904).

Variables	Mobile phone ownership			
	Males (n=305)		Females (n=599)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)				
18-30	—[a]	—	1.89 (0.9-3.59)	.05
31-40	—	—	2.42 (1.24-4.70)	.01
41-50	—	—	2.16 (1.01-4.59)	.05
50+[b]	—	—	—	—
Education				
No school	0.51 (0.08-3.46)	.49	0.714 (0.213-2.394)	.58
Incomplete school	2.16 (0.32-14.49)	.45	1.219 (0.369-4.031)	.75
High school diploma	0.76 (0.09-6.05)	.79	1.34 (0.319-5.619)	.69
Some college or college graduate[b]	—	—	—	—
Household education				
No school	0.54 (0.11-2.72)	.46	0.27 (0.12-0.59)	.001
Incomplete school	0.52 (0.13-2.13)	.36	0.54 (0.27-1.06)	.07
High school diploma	1.29 (0.27-6.14)	.75	0.56 (0.26-1.19)	.13
Some college or college graduate[b]	—	—	—	—
Type of family				
Broken[b]	—	—	—	—
Extended	—	—	0.6 (0.10-3.93)	.63
Joint	—	—	1.39 (0.29-6.78)	.68
Nuclear	—	—	1.38 (0.29-6.64)	.68
Total earning members in the household				
No earning member	0.06 (0.01-0.54)	.01	—	—
1 earning member	0.98 (0.32-2.99)	.97	—	—
2 earning members	1.09 (0.32-3.69)	.89	—	—
3 or more earning members[b]	—	—	—	—
Housing type				
Nonconcrete[b]	—	—	—	—
Concrete	0.41 (0.14-1.17)	.09	1.37 (0.71-2.62)	.35
Semiconcrete	2.05 (0.86-4.90)	.11	0.77 (0.40-1.48)	.43
Type of toilet facility				
In-house[b]	—	—	—	—
Public place	2.23 (0.72-6.93)	.16	1.59 (1.04-2.43)	.03
Open defecation	2.88 (1.09-7.63)	.03	0.47 (0.26-0.87)	.02
Television ownership				
No[b]	—	—	—	—
Yes	1.27 (0.34-4.81)	.72	1.06 (0.54-2.07)	.87
Television ownership with satellite television service				
No[b]	—	—	—	—

Variables	Mobile phone ownership			
	Males (n=305)		Females (n=599)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Yes	3.30 (0.93-11.69)	.06	1.28 (0.71-2.31)	.40
Smoking				
No[b]	—	—	—	—
Yes	0.61 (0.30-1.22)	.16	—	—

[a] Empty cells indicate that the variables were not significant in the bivariate analysis for the respective gender group.

[b] Reference group.

Multivariable Analysis of Factors Associated With Internet Access Within Males and Females

Age of the study participant (*P*=.04), education (*P*=.01), and household education (*P*<.001) remained significantly associated with mobile phone internet access among males (Table 2). Males between the ages of 18 and 30 years had higher odds of internet access (OR 2.60, 95% CI 1.02-6.67; *P*=.04). The odds of internet access were lower among males with no schooling (OR 0.09, 95% CI 0.02-0.57; *P*=.01) or incomplete schooling (OR 0.16, 95% CI 0.03-0.83; *P*=.03). Living in households where the highest educational level was at most a high school diploma was associated with a lower odds of internet access (OR 0.29, 95% CI 0.09-0.89; *P*=.03).

Household educational level (*P*<.001) and housing type (*P*=.03) remained significantly associated with mobile phone internet access among females. Living in households where the highest educational level was less than a high school diploma was associated with a lower odds of internet access (OR 0.16, 95% CI: 0.09-0.31; *P*<.001). In addition, the odds of internet access were lower among females living in houses made of semiconcrete (OR 0.38, 95% CI 0.16-0.92).

Table 2. Multivariable analysis showing predictors of internet access within the male and female study participants (N=220).

Variables	Internet access			
	Males (n=305)		Females (n=599)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)				
18-30	2.60 (1.02-6.67)	.04	—[a]	—
31-40	1.57 (0.56-4.41)	.39	—	—
41-50	0.86 (0.28-2.69)	.79	—	—
50+[b]	—	—	—	—
Education				
No school	0.09 (0.02-0.57)	.01	0.65 (0.22-1.93)	.44
Incomplete school	0.16 (0.03-0.83)	.03	1.82 (0.63-5.28)	.27
High school diploma	0.34 (0.06-2.02)	.24	1.12 (0.29-4.29)	.86
Some college or college graduate[b]	—	—	—	—
Household education				
No school	0.26 (0.05-1.28)	.09	0.16 (0.06-0.42)	<.001
Incomplete school	0.23 (0.08-0.61)	.004	0.16 (0.09-0.31)	<.001
High school diploma	0.29 (0.09-0.89)	.03	0.68 (0.34-1.33)	.26
Some college or college graduate[b]	—	—	—	—
Type of family				
Broken[b]	—	—	—	—
Extended	—	—	1.27 (0.11-14.50)	.85
Joint	—	—	1.98 (0.21-18.69)	.55
Nuclear	—	—	0.80 (0.09-7.47)	.85
Total earning members in the household				
No earning member	0.17 (0.01-2.22)	.18	—	—
1 earning member	0.45 (0.18-1.14)	.09	—	—
2 earning members	1.29 (0.50-3.34)	.59	—	—
3 or more earning members[b]	—	—	—	—
Housing type				
Nonconcrete[b]	—	—	—	—
Concrete	2.51 (0.61-10.36)	.20	0.79 (0.35-1.79)	.57
Semiconcrete	2.62 (0.64-10.74)	.18	0.38 (0.16-0.92)	.03
Type of toilet facility				
In-house[b]	—	—	—	—
Public place	1.94 (0.90-4.16)	.09	0.70 (0.42-1.17)	.17
Open defecation	—	—	1.25 (0.53-2.92)	.61
Television ownership				
No[b]	—	—	—	—
Yes	—	—	1.94 (0.72-5.23)	.19
Television ownership with satellite television service				
No[b]	—	—	—	—

Variables	Internet access			
	Males (n=305)		Females (n=599)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Yes	—	—	1.40 (0.67-2.95)	.37

[a] Empty cells indicate that the variables were not significant in the bivariate analysis for the respective gender group.

[b] Reference group.

Multivariable Analysis of Factors Associated With Knowledge of SMS Text Messaging Within Males and Females

Household educational level (*P*<.001), number of earning members (*P*=.01), and housing type (*P*=.02) remained significantly associated with the knowledge of text messaging among males (Table 3). Males from households who had no schooling (OR 0.05, 95% CI 0.02-0.18) or less than high schooling (OR 0.23, 95% CI 0.09-0.54) had a lower odds of text messaging. Having no earning member in the household was also associated with a lower odds of text messaging among males (OR 0.04, 95% CI 0.004-0.43). Males who lived in houses made of semiconcrete had higher odds of text messaging (OR 3.94, 95% CI 1.28-12.09).

Individual education (*P*=.04), household education (*P*<.001), satellite television service (*P*=.03), and smoking (*P*=.01) remained significantly associated with text messaging. Females who had no schooling (OR 0.24, 95% CI 0.06-0.97) or incomplete schooling (OR 0.23, 95% CI 0.12-0.46) had a lower odds of text messaging. Females who owned a television with satellite service had higher odds of text messaging (OR 2.06, 95% CI 1.08-3.93). Smoking behavior was associated with lower odds of text messaging (OR 0.45, 95% CI 0.26, 0.79).

Table 3. Multivariable analysis showing predictors of text messaging within the male and female study participants (N=904).

Variables	Text messaging			
	Males (n=305)		Females (n=599)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)				
18-30	—[a]	—	1.81 (0.86-3.81)	.12
31-40	—	—	1.95 (0.92-4.16)	.08
41-50	—	—	2.18 (0.92-5.15)	.08
50+[b]	—	—	—	—
Education				
No school	—	—	0.24 (0.06-0.97)	.04
Incomplete school	—	—	0.57 (0.15-2.24)	.42
High school diploma	—	—	1.37 (0.26-7.12)	.71
Some College or college graduate[b]	—	—	—	—
Household education				
No school	0.05 (0.02-0.18)	<.001	0.10 (0.04-0.25)	<.001
Incomplete school	0.23 (0.09-0.54)	.001	0.23 (0.12-0.46)	<.001
High school diploma	0.47 (0.17-1.28)	.14	0.47 (0.21-1.02)	.06
Some College or college graduate[b]	—	—	—	—
Type of family				
Broken[b]	—	—	—	—
Extended	—	—	0.57 (0.07-4.48)	.59
Joint	—	—	1.81 (0.30-10.89)	.52
Nuclear	—	—	0.80 (0.13-4.82)	.81
Total earning members in the household				
No earning member	0.04 (0.004-0.43)	.01	—	—
1 earning member	0.47 (0.19-1.16)	.10	—	—
2 earning members	0.77(0.29-2.03)	.59	—	—
3 or more earning members[b]	—	—	—	—
Housing type				
Nonconcrete[b]	—	—	—	—
Concrete	2.99 (0.99-9.03)	.05	1.66 (0.81-3.43)	.17
Semiconcrete	3.94 (1.28-12.09)	.02	0.50 (0.24-1.06)	.07
Type of toilet facility				
In-house[b]	—	—	—	—
Public place	0.90 (0.48-1.68)	.74	1.14 (0.73-1.79)	.56
Open defecation	0.42 (0.15-1.23)	.11	0.79 (0.39-1.62)	.53
Television ownership				
No[b]	—	—	—	—
Yes	1.81 (0.52-6.34)	.35	1.29 (0.59-2.82)	.52
Television ownership with satellite television service				
No[b]	—	—	—	—

Variables	Text messaging			
	Males (n=305)		Females (n=599)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Yes	2.01 (0.66-6.11)	.22	2.06 (1.08-3.93)	.03
Smoking				
No[b]	—	—	—	—
Yes	—	—	0.45 (0.26-0.79)	.01

[a] Empty cells indicate that the variables were not significant in the bivariate analysis for the respective gender group.

[b] Reference group.

Discussion

Principal Findings

The results of this study showed that more than half of the study participants owned a mobile phone (602/904, 66.3%), 49.3% (446/904) of them were familiar with sending text messages, and 24.3% (220/904) of them had internet service on their mobile phones. Variables associated with mobile phone ownership included gender, age, household educational level, number of earning members, housing type, and type of toilet facility. Variables associated with text messaging included the age of the study participant, individual education, household education, number of earning members in the household, housing type, and satellite television service. Variables significantly associated with internet access included gender, age, and household education. High-risk behaviors, including smoking and alcohol consumption were not significantly associated with mobile phone ownership, internet access, or text messaging.

This study identified important differences in the demographic and behavioral correlates of mobile phone ownership, access to the internet, and text messaging between males and females. Females were half as likely to own mobile phones compared with males (OR 0.53, 95% CI 0.37-0.76), less likely to have internet access (OR 0.79, 95% CI 0.56-1.11), or knew how to send text messages (OR 0.93, 95% CI 0.66-1.31). This finding is consistent with several similar studies in the literature [12]. Prior studies have shown that women in low- and middle-income countries (LMICs) are 14% less likely to own mobile phones compared with men. In addition, internet access is 12% lower among women compared with men [13].

Variables associated with mobile phone ownership between males and females included age, individual education, and housing type, in the adjusted analysis ($P<.05$). Being older than 30 years of age (OR 2.19, 95% CI 1.35-3.58), having no education (OR 0.31, 95% CI 0.14-0.66), and living in houses made of concrete (OR 2.49, 95% CI 1.50-4.14) were associated with mobile phone ownership. Some of these findings were not consistent with prior literature [3]. A prior study assessing determinants of household phone ownership in rural Bangladesh showed higher odds of mobile phone ownership among younger participants between the ages of 20 and 24 (OR 1.22, 95% CI 1.03-1.44), with lesser odds among participants aged 30 years or more (OR 0.95, 95% CI 0.77-1.18) [3]. A consistent finding of the highlighted literature with this study was the significance of wealth index measures (such as living in houses made of concrete) as a predictor of mobile phone ownership ($P<.001$). In particular, the highest quartile of wealth index was a key predictor of mobile phone ownership in rural Bangladesh [3].

In addition to age and individual education, the number of earning members in a household and housing type were predictors of text messaging between males and females in the adjusted analysis. Having 1 earning member or less was associated with reduced odds of text messaging (OR 0.547, 95% CI 0.329-0.909). This finding is consistent with prior literature in similar settings [14]. Such findings indicate that a high financial debt (especially in the absence of male family members), coupled with family responsibilities that affect finances (such as having several dependent family members), are possible mechanisms that explain the lower technology utilization among females [14].

Individual education was the sole predictor of internet access between males and females in the adjusted analysis. Having less than a high school education was significantly associated with reduced odds of internet access between males and females (OR 0.34, 95% CI 0.15-0.75). This finding was consistent with prior studies highlighting the role of literacy in internet utilization across LMICs. According to a qualitative analysis of socioeconomic correlates of the gender digital divide in Rwanda, a male participant stated as follows [15]:

> *For most Rwandan women, particular barriers are illiteracy, lack of familiarity with the main languages of computer technology and the internet, lack of operational training in computers, heavy household tasks, and lack of self-confidence* [male, 50 years]

The relevance of individual education as a key predictor of all three measures of technology used between males and females in this study is consistent with prior literature [12,14]. These findings have been attributed to patriarchal norms, which are arguably more prevalent in the South Asian context [12]. One of such patriarchal norms is the dedication of funds for educating male children, but rather preparing for the wedding of female children [14]. Such norms largely promote mobile phone ownership and technology usage among male family members, although limiting their use among women, who would often require permission to use mobile phones [16].

Our study findings comparing the differences in technology outcomes between gender groups highlighted disparate predictors of mobile phone ownership, internet access, and text messaging between males and females. For instance, individual education was a predominant factor associated with all three technology outcomes. Age and housing type were only predictive of mobile phone ownership and text messaging, and the number of earning members was only predictive of text messaging. These disparate findings across these predictors suggested that individual differences within gender groups may also be evident across the technology outcomes.

Subsequent investigations indicated that predictors of the technology outcomes differed significantly within males and females. These findings are suggestive of disparate levels of social class that are predominant even with slum settings, consistent with prior literature [17]. For instance, slum studies conducted in the Indian states of Chennai showed that individuals residing on the outskirts or margins of slums that adjoin wealthy neighborhoods are likely to absorb the culture of these wealthy Chennai neighborhoods. Owing to this, they were found to have higher levels of technology awareness and utilization [17]. It is also likely that such slum residences that are situated closer to wealthy neighborhoods may have lower crime and poverty rates, which are major predictors of the disparate access to technology by gender [16]. These findings can be extrapolated to this study in explaining the individual differences within gender categories with respect to the technology outcomes studied, as discussed below.

Our results showed that the type of toilet facility in households was significantly associated with mobile phone ownership. However, this association was different among males and females. Using a public toilet facility was not significantly associated with mobile phone ownership among males but was significant for females. In addition, open defecation was significantly associated with mobile phone ownership among females and males. These findings can be explained by prior literature indicating that open defecation is a marker of reduced SES [18]. In comparison to public places where slum residents have to pay around US $0.03 to US $0.04 to use a community public toilet, individuals practicing open defecation do not have to pay for this practice. The finding of open defecation being protective among females could be attributed to a higher sense of insecurity among females with using shared public latrines [19]. In addition, prior literature has indicated that although policies have been enacted to foster the provision of gender-specific toilets and infrastructure, females remain disproportionately affected by a lack of female-specific sanitation facilities compared with their male counterparts. In particular, findings have recorded a 66% disparity between toilets for men and women as of April 2019 [19]. Taken together, open defecation as a marker of reduced socio-economic status ultimately impacts female ownership of household assets and technologies [19].

Our study also showed that among females, household education was a predictor of all the technology outcomes, in the adjusted analysis. The finding that household education (and not individual education) was a key predictor of all the technology outcomes within females is consistent with the dominance of patriarchal norms within slum settings, which serve to marginalize women. In particular, gender inequities have been attributed to norms such as lack of prioritization of women's education from the childhood stage, women being forced to get married before they can attain financial independence, family responsibilities preventing women from working outside their homes, or in-laws forcing women to work in low-profit family businesses rather than letting them earn independently, and much more [14]. These combined factors serve to increase women's financial dependence on the men within their households; thus, the educational level of these women become less significant, as that of the household earning member takes precedence [14].

This in turn lends support to our study finding that having 1 earning member or less was a key predictor of both internet access and text messaging (corresponding to higher levels of technology usage) among men. Our study findings showed that among males, the number of earning members was a predictor of both mobile phone ownership and text messaging. This situation may be explained by the likelihood for women to depend more on the men in the households for financial sustenance; hence, the lack of a household earning member may significantly impair the financial status of the households. This phenomenon may then translate into a reduced ownership of household infrastructure and assets such as housing type (concrete vs semiconcrete vs nonconcrete) and television ownership, which are essentially markers of wealth, with higher values indicating better SES.

Strengths and Limitations

A limitation of this study is the convenience approach employed in the identification of slums and, ultimately, the households that were interviewed. This approach may have introduced selection bias and affected the generalizability of the results to the entire population across Indian slums. This study is, however, generalizable to individuals residing in the unnotified slums included in this study. This study had several strengths. Our study findings provide an in-depth exploration of individual gender differences in the digital divide and highlight relevant measures of these differences. Although our study does not demonstrate any causal or quasi-causal claims, it highlights the possible areas of intervention that are in line with the identified predictors.

Conclusions

Our study findings show disparate access to the technology outcomes *within* males and females in slum settings and lend support for further research to examine the causal mechanisms promoting these differences. Such mechanisms may proffer significant solutions to address the technology divide within gender groups and ultimately between gender groups. Specifically, our study findings suggest that improving household education is crucial to address the disparate access and utilization of mobile phones, internet, and text messaging among women in slum settings, owing to the consistency in household educational level as a predictor across all these technology indicators. In addition, the mechanisms by which the number of household earning members influences the disparate access to technology among men, call for further exploration. Finally, although our chosen study focus was on gender disparities in access to specific technology outcomes, future studies could explore the impact of the gender digital divide on access to health care and other health-related behaviors beyond those covered in this study. Internet and mobile phone usage are required for electronic health and mobile health technologies to promote the utilization of health care services.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variables associated with mobile phone ownership, internet access, and text messaging among the study participants.
[DOCX File , 22 KB - jmir_v22i6e14714_app1.docx]

Multimedia Appendix 2
Predictors of mobile phone ownership, internet access, and text messaging among the study participants.
[DOCX File , 21 KB - jmir_v22i6e14714_app2.docx]

Multimedia Appendix 3
Differences in mobile phone ownership across gender categories.
[DOCX File , 16 KB - jmir_v22i6e14714_app3.docx]

Multimedia Appendix 4
Differences in internet access across gender categories.
[DOCX File , 15 KB - jmir_v22i6e14714_app4.docx]

Multimedia Appendix 5
Differences in text messaging across gender categories.
[DOCX File , 15 KB - jmir_v22i6e14714_app5.docx]

Multimedia Appendix 6
Predictors of mobile phone ownership, internet access, and text messaging between males and females.
[DOCX File , 18 KB - jmir_v22i6e14714_app6.docx]

Multimedia Appendix 7
Within-gender variation related to mobile phone ownership.
[DOCX File , 17 KB - jmir_v22i6e14714_app7.docx]

Multimedia Appendix 8
Within-gender variation related to internet access.
[DOCX File , 22 KB - jmir_v22i6e14714_app8.docx]

Multimedia Appendix 9
Text messaging within the male and female study participants.
[DOCX File , 22 KB - jmir_v22i6e14714_app9.docx]

References

1. Chipeva P, Cruz-Jesus F, Oliveira T, Irani Z. Digital divide at individual level: evidence for eastern and western European countries. Gov Inf Q 2018 Sep;35(3):460-479 [FREE Full text] [doi: 10.1016/j.giq.2018.06.003]
2. Han C. South African perspectives on mobile phones: challenging the optimistic narrative of mobiles for development. Int J Commun 2012;6:25 [FREE Full text]
3. Tran MC, Labrique AB, Mehra S, Ali H, Shaikh S, Mitra M, et al. Analyzing the mobile 'digital divide': changing determinants of household phone ownership over time in rural Bangladesh. JMIR Mhealth Uhealth 2015 Feb 25;3(1):e24 [FREE Full text] [doi: 10.2196/mhealth.3663] [Medline: 25720457]
4. Hong YA, Zhou Z, Fang Y, Shi L. The digital divide and health disparities in China: evidence from a national survey and policy implications. J Med Internet Res 2017 Sep 11;19(9):e317 [FREE Full text] [doi: 10.2196/jmir.7786] [Medline: 28893724]

5. van Dijk JA. Digital divide: impact of access. In: The International Encyclopedia of Media Effects. Hoboken, New Jersey: Wiley-Blackwell; 2017:1-11.
6. United Nations. 2019. Goal 5: Achieve Gender Equality and Empower All Women and Girls URL: https://www.un.org/sustainabledevelopment/gender-equality/ [accessed 2019-05-14] [WebCite Cache ID 78MfICp3K]
7. Singh S. Bridging the gender digital divide in developing countries. J Child Media 2017 Mar 24;11(2):245-247 [FREE Full text] [doi: 10.1080/17482798.2017.1305604]
8. Liang T. Matrix analysis of the digital divide in ehealth services using awareness, want, and adoption gap. J Med Internet Res 2012 Feb 13;14(1):e11 [FREE Full text] [doi: 10.2196/jmir.1670] [Medline: 22329958]
9. Un-Habitat. Planning and Design for Sustainable Urban Mobility: Global Report on Human Settlements 2013. Abingdon, United Kingdom: Routledge; 2013.
10. Chakravarty S, Dand S. Research Papers in Economics. Indian Institute of Management Ahmedabad, Research and Publication Department URL: https://ideas.repec.org/s/iim/iimawp.html [accessed 2019-07-11]
11. Joshi A, Arora A, Amadi-Mgbenka C, Mittal N, Sharma S, Malhotra B, et al. Burden of household food insecurity in urban slum settings. PLoS One 2019;14(4):e0214461 [FREE Full text] [doi: 10.1371/journal.pone.0214461] [Medline: 30939157]
12. Bala S, Singhal P. Gender digital divide in India: a case of inter-regional analysis of Uttar Pradesh. J Inf Commun Eth Soc 2018 Apr 13;16:173-192 [FREE Full text] [doi: 10.1108/jices-07-2017-0046]
13. Herbert S. Assets Publishing. 2017. Digital Development and the Digital Gender Gap URL: https://assets.publishing.service.gov.uk/media/5a5f228f40f0b652634c6f4a/249-Digital-development-and-the-digital-gender-gap.pdf [accessed 2019-07-14] [WebCite Cache ID 78Mfgew6h]
14. Potnis D. Inequalities creating economic barriers to owning mobile phones in India. Inf Dev 2016 Jul 8;32(5):1332-1342 [FREE Full text] [doi: 10.1177/0266666915605163]
15. Mumporeze N, Prieler M. Gender digital divide in Rwanda: a qualitative analysis of socioeconomic factors. Telemat Inform 2017 Nov;34(7):1285-1293 [FREE Full text] [doi: 10.1016/j.tele.2017.05.014]
16. Rashid A. Digital inclusion and social inequality: gender differences in ICT access and use in five developing countries. Gend Technol Dev 2017 Oct 25;20(3):306-332 [FREE Full text] [doi: 10.1177/0971852416660651]
17. Rangaswamy N, Arora P. The mobile internet in the wild and every day: digital leisure in the slums of urban India. Int J Cult Stud 2016 Jul 24;19(6):611-626 [FREE Full text] [doi: 10.1177/1367877915576538]
18. Heijnen M, Routray P, Torondel B, Clasen T. Shared sanitation versus individual household latrines in urban slums: a cross-sectional study in Orissa, India. Am J Trop Med Hyg 2015 Aug;93(2):263-268 [FREE Full text] [doi: 10.4269/ajtmh.14-0812] [Medline: 26123953]
19. Kopikar S. The Guardian. 2017. Death-trap Toilets: The Hidden Dangers of Mumbai's Poorest Slums URL: https://www.theguardian.com/global-development-professionals-network/2017/feb/27/death-trap-toilets-mumbai-india-slums [accessed 2019-07-11]

Abbreviations

ICT: Information and Communication Technology
IRB: Institutional Review Board
ITU: International Telecommunication Union
LMICs: low- and middle-income countries
OR: odds ratio
SES: socioeconomic status
UN: United Nations

Edited by G Eysenbach; submitted 14.05.19; peer-reviewed by L Madrigal, J op den Buijs, F Yamin; comments to author 03.10.19; revised version received 18.10.19; accepted 16.12.19; published 22.06.20.

Please cite as:
Joshi A, Malhotra B, Amadi C, Loomba M, Misra A, Sharma S, Arora A, Amatya J
Gender and the Digital Divide Across Urban Slums of New Delhi, India: Cross-Sectional Study
J Med Internet Res 2020;22(6):e14714
URL: https://www.jmir.org/2020/6/e14714
doi:10.2196/14714
PMID:32343670

Original Paper

Internet and Health Information Technology Use and Psychological Distress Among Older Adults With Self-Reported Vision Impairment: Case-Control Study

Namkee G Choi[1], PhD; Diana M DiNitto[1], PhD; Othelia EunKyoung Lee[2], PhD; Bryan Y Choi[3], MD

[1]Steve Hicks School of Social Work, The University of Texas at Austin, Austin, TX, United States
[2]School of Social Work, University of North Carolina at Charlotte, Charlotte, NC, United States
[3]Department of Emergency Medicine, Brown University, Providence, RI, United States

Corresponding Author:
Namkee G Choi, PhD
Steve Hicks School of Social Work
The University of Texas at Austin
1925 San Jacinto Blvd
Austin, TX, 78712
United States
Phone: 1 5122329590
Email: nchoi@austin.utexas.edu

Abstract

Background: The number of older adults with vision impairment (VI) is growing. As health care services increasingly call for patients to use technology, it is important to examine internet/health information technology (HIT) use among older adults with VI.

Objective: This study aimed to examine (1) the rates of internet/HIT use among older adults with VI compared with a matched sample of their peers without VI, (2) associations of VI with internet/HIT use, and (3) association of HIT use with psychological distress, assessed with the Kessler-6 screen.

Methods: Data were obtained from the 2013 to 2018 US National Health Interview Survey. Older adults (aged ≥65 years) with self-reported VI were matched with older adults without VI, in a 1:1 ratio, based on age, sex, number of chronic medical conditions, and functional limitations (N=2866). Descriptive statistics and multivariable logistic regression models, with sociodemographic factors, health conditions, health insurance type, and health care service use as covariates, were used to examine the research questions.

Results: In total, 3.28% of older adults (compared with 0.84% of those aged 18-64 years) reported VI, and 25.7% of them were aged ≥85 years. Those with VI were significantly more socioeconomically disadvantaged than those without VI and less likely to use the internet (adjusted odds ratio [aOR] 0.64, 95% CI0.49-0.83) and HIT (aOR 0.74, 95% CI 0.56-0.97). However, among internet users, VI was not associated with HIT use. HIT use was associated with lower odds of mild/moderate or serious psychological distress (aOR 0.62, 95% CI 0.43-0.90), whereas VI was associated with greater odds of mild/moderate or serious distress (aOR 1.84, 95% CI 1.36-2.49). Health care provider contacts were also associated with higher odds of internet or HIT use.

Conclusions: Compared with their matched age peers without VI, older adults with VI are less likely to use HIT because they are less likely to use the internet. Socioeconomically disadvantaged older adults experiencing a digital divide need help to access information and communication technologies through a fee waiver or subsidy to cover internet equipment and subscription and ensure continuous connectivity. Older adults with VI who do not know how to use the internet/HIT but want to learn should be provided instruction, with special attention to accessibility features and adaptive devices. Older adults with a low income also need better access to preventive eye care and treatment of VI as well as other health care services.

(J Med Internet Res 2020;22(6):e17294) doi:10.2196/17294

KEYWORDS

older adults; vision impairment; HIT; psychological distress; digital divide; mobile phone

Introduction

Background

A significant portion of older adults in the United States have vision impairment (VI) and associated disability. According to the 2014 American Community Survey, 4.3% (1.1 million) of the civilian, noninstitutionalized population of adults aged 65 to 74 years and 10% (1.9 million) of those aged ≥75 years reported blindness or serious trouble seeing, even when wearing glasses [1]. In the 2014 National Health Interview Survey (NHIS), 13.5% (6.1 million noninstitutionalized adult participants aged ≥65 years, including 350,000 who were blind) reported “any trouble seeing, even when wearing glasses or contact lenses” [1]. VI increases with advancing age, and the rate for the ≥85 age group in the 2014 NHIS was 22.3% [1]. The Centers for Disease Control and Prevention’s (CDC) analysis of the 2016 Behavioral Risk Factor Surveillance System (BRFSS) data also found that 6.6% of noninstitutionalized older adults reported having a vision-related disability, compared with 2.7% in the 18 to 44 age group and 6.1% in the 45 to 64 age group [2]. The BRFSS data showed that regardless of age group, VI (similar to other types of disability) is more prevalent among women than men, racial/ethnic minorities than non-Hispanic whites, and among those with lower income.

VI in older adults is attributable to refractive error (correctable with glasses) and eye disease (often not correctable or requiring surgical/medical interventions: age-related macular degeneration, ocular complications of diabetes mellitus, glaucoma, and age-related cataracts) [3,4]. According to Swenor et al [3], in the United States, uncorrected refractive error accounts for approximately 79% of VI, and cataracts are the most common correctable eye disease, accounting for approximately 50% of VI from eye disease. There are also effective strategies to prevent or delay late-life vision loss and blindness resulting from other defined diseases [4].

VI has a variety of functional consequences, as it can restrict certain activities that require good vision. One such activity may be technology use, including internet and health information technology (HIT) use. The rate of internet use in older adults has steadily risen over the past decade, with 67% using the internet and 51% having a home broadband connection in 2017 [5]. According to the Health Information National Trends Survey data, seeking health information and communicating with physicians on the web were the most significant areas of increase in internet use among older adults between 2003 and 2011/2012 [6]. However, although the digital health technology divide between genders, racial/ethnic groups, rural/urban residents, and those of various health statuses has narrowed over the years, HIT use has remained lower among those in advanced age, racial/ethnic minorities, and those with less education and lower income [6-8]. Lower levels of internet use among disabled older adults with a low income are largely attributable to the lack of financial resources to obtain computers and technology or medical conditions, disabilities, and associated pain that restrict use [9]. Keränen et al [10] also found that physical frailty was associated with lower information and communications technology (ICT) use independent of age, education, and opinions of technology use among older adults. However, studies on the internet and HIT use among older adults in the United States with VI compared with those without VI are not available, indicating a major gap in knowledge regarding VI’s impact on the digital divide.

As health care services, including provider-patient communications, increasingly rely on technology, the use of HIT by older adults has significant implications for the health care system and older adults themselves. HIT may reduce health care spending by providing preventive health promotion information, facilitating communication with health care providers, and improving health care quality and outcomes for older adults [8,11]. Longitudinal data show that HIT use may be associated with fewer physician visits among older adults with certain chronic health conditions (eg, diabetes). However, cross-sectional data show that HIT use is associated with higher health service use (eg, general practitioners, medical specialists, eye doctors, physical therapists/occupational therapists [PT/OT]) [7,12]. The latter may also be positive, as health information seeking and other HIT use behaviors, either encouraged by providers or initiated by the patient, can contribute to increased disease-related knowledge and better adherence to prevention and treatment regimens. Increased HIT use thus has significant potential for health care cost savings and improved health care quality and outcomes, especially among socioeconomically disadvantaged older adults and those with chronic medical conditions and disability.

ICT use may also impact how older adults perceive social support, loneliness, depression, psychological well-being, and quality of life, but findings vary. Some studies show that internet use results in higher levels of social support, life satisfaction and well-being, and reduced loneliness and depression [13-16]. Other studies indicate that variables such as living arrangements, ethnicity, and contact with family mediate these positive relationships and that more frequent use of ICT was associated with more psychological distress and less sense of community among older adults who were lonelier [17-19]. Elliot et al [20] found that although ICT use was not directly related to depressive symptoms or well-being, it acted as a moderator. In essence, functional limitation was a stronger predictor of depressive symptoms for high users, and ill health was a stronger predictor for nonusers or limited users. However, to our knowledge, there is no published study on the association of internet/HIT use with psychological well-being among older adults with VI.

Study Aims

Given the increasing number of older adults with VI, this case-control study examined (1) rates of internet and HIT use among a nationally representative sample of older adults with VI, compared with a matched sample of their peers without VI, (2) associations between VI and internet and HIT use, and (3) association of HIT use with psychological distress. The study hypotheses were that VI will be associated with lower odds of internet/HIT use (H1), and HIT use will be associated with lower odds of mild/moderate or serious psychological distress, controlling for VI (H2). Other covariates for the multivariable hypotheses testing were sociodemographic factors, health conditions, health insurance type, and health care service use. These findings may help identify VI as a barrier to internet/HIT use among older adults and suggest ways to improve access to technology for older adults with VI.

Methods

Data and Sample

Data from the 2013 to 2018 NHIS public-use data files were downloaded from the CDC's National Center for Health Statistics website. The NHIS is an annual, cross-sectional household survey that is the principal source of information on the health and health care access of the civilian, noninstitutionalized US population [21]. For each sampled household, interviews are conducted (mostly face-to-face) with an adult family member who answers questions about each family member's demographic and health status characteristics. The NHIS also collects more detailed health and other data from 1 sample adult from each household, which is the primary data source for this study. Combining 6 consecutive years of annual NHIS data resulted in 190,113 sample adult respondents aged 18 to ≥85 years (NHIS public-use data sets do not provide the chronological age of those aged >85 years). Of the 190,113 sample adults, 19.4% (n=48,287) were aged ≥65 years. In this study, we focused on those aged ≥65 years with VI and their matched peers without VI (n=2866).

Measures

Vision Impairment

In the NHIS, respondents were asked about vision ("any trouble seeing, even when wearing glasses or contact lenses" and "blind or unable to see at all"). They were also asked a series of questions about the levels of difficulty (not at all to cannot do it at all) performing physical functions (eg, walking, climbing) and social activities (eg, going out to social events, participating in social activities) without special equipment. Those who reported any difficulty in any of these activities/functions were then asked a series of follow-up questions about whether the difficulty was caused by specific conditions (eg, vision problems, illnesses, injuries). In this study, we refer to older adults who reported that "vision/problem seeing causes difficulty with activities" as those with VI.

Internet Use

Respondents were asked if they had used the internet in the past 12 months (yes=1 and no=0).

Health Information Technology Use

Respondents were asked if they had (1) looked up health information on the internet, (2) filled a prescription on the internet, (3) scheduled a medical appointment on the internet, and (4) communicated with a health care provider by email in the past 12 months. In this study, HIT use refers to any of these 4 activities.

Psychological Distress

This was assessed using the Kessler-6 screen (K6), a global measure of distress that includes depressive- and anxiety-related symptomology over the 4-week period before test administration [22]. The 6 items were how often you felt nervous, restless/fidgety, so depressed that nothing cheered you up, hopeless, worthless, and that everything was an effort (0=none of the time, 1=little of the time, 2=some of the time, 3=most of the time, and 4=all of the time). K6 has reported a sensitivity of 0.36, a specificity of 0.96, and a total classification accuracy of 0.92 in predicting severe mental illness, defined as any Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) disorder, other than a substance use disorder, with a Global Assessment of Functioning score less than 60 occurring in the past 12 months [22]. K6 scores of 13 to 24 (ie, serious distress) indicate a probable DSM-IV disorder and significant impairment in functioning, and K6 scores of 8 to 12 (mild/moderate distress) indicate a probable diagnosable mental illness but with less severe impairment in functioning [23]. Cronbach α for the 6 items among the study sample was 0.86.

Sociodemographic Factors

Sociodemographic factors are as follows: (1) age (65-74, 75-84, and ≥85 years), (2) gender, (3) race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, Asian, and other), (4) marital status (not married vs married), (5) education (bachelor's degree vs no degree), and (6) ratio of family income to the official US poverty threshold (<200%, 200%-399%, ≥400%, and missing).

Health Status

Indicators of health status are as follows: (1) number (0-10) of diagnosed chronic medical conditions (eg, arthritis, asthma, diabetes, hypertension, heart disease, stroke, chronic kidney disease, liver disease, lung disease, and cancer) with which the respondent had ever been diagnosed, and (2) hearing impairment ("hearing problem causes difficulty with activity," yes=1 and no=0).

Health Insurance

Respondents were asked if they had Medicare, Medicaid, private health insurance, and veterans/military insurance coverage (yes=1 and no=0 for each) in the past 12 months.

Health Care Use

Respondents were asked if they had in the past 12 months (1) had an overnight hospitalization, (2) received health care 10 or more times, and (3) saw/talked to a general practitioner, eye doctor, medical specialist, PT/OT, or mental health professional (yes=1 and no=0 for each).

Data Analysis

All analyses were conducted with Stata 15 (Stata Corp)/MP's svy function to account for the NHIS's stratified, multistage probability sampling design and to ensure that variance estimates incorporate the full sampling design. All statistics were weighted except for sample sizes. First, using the ccmatch function [24] in Stata, we matched older adults with VI with those without VI in a 1:1 ratio, based on age, sex, number of chronic medical conditions, any other functional limitation than VI, and self-response vs proxy response status. Second, we used chi-square and 2-tailed t tests to describe and compare older adults with and without VI on sociodemographic and health statuses, K6 scores, health insurance, and health care use. For

K6 scores, we excluded 57 respondents with missing data on 4 or more items but included 6 respondents who had missing data on 1 item and 3 respondents who had missing data on 2 or 3 items after replacing the missing data with the mean of their nonmissing items on the scale for the summed score. Third, we used chi-square tests and 95% CIs to compare older adults with and without VI on their internet and HIT use. Fourth, we tested H1 (association of VI with internet/HIT use) with the study sample using 2 logistic regression models with VI (vs no VI) as the independent variable and internet use (vs nonuse) and any HIT use (vs nonuse) as the dependent variables. We also used logistic regression to examine the association of VI with any HIT use among internet users. Finally, we tested H2 (association of HIT use with psychological distress) using a logistic regression model with any HIT use (vs nonuse) as the independent variable and mild/moderate or severe distress versus no distress (ie, K6 scores 8-24 vs K6 scores 0-7) as the dependent variable. Variance inflation factor diagnostics, using a cutoff of 2.50 [25], showed that multicollinearity among covariates was not a concern. Logistic regression results are presented as adjusted odds ratios (aORs) with 95% CI. Statistical significance was set at a *P*<.05.

Results

Study Sample Selection

Of the 48,287 sample adult respondents aged ≥65 years in the 2013 to 2018 NHIS, 16.07% (n=7896, including 398 blind respondents) reported that they had “any trouble seeing, even when wearing glasses or contact lenses.” However, only 18.73% of those who were not blind but had vision problems and 53.15% of blind older adults reported vision-caused difficulty with activities. As a result, 1630 older adults, including 210 blind persons or 3.28% of all NHIS sample adults aged ≥65 years, were deemed to have VI. The 3.28% rate of VI was significantly higher than the 0.84% rate of VI among the NHIS sample adults aged 18 to 64 years. Annual rates of “any trouble seeing” among the age group ≥65 years showed an overall increasing trend over the study period; however, annual VI rates among the age group ≥65 years were not significantly different over the study period.

Of the 1630 older adults with VI, 1598 were matched in a 1:1 ratio with 1598 older adults without VI. After excluding 165 from each group whose interviews were done by proxy (as some questions were not asked of the proxy respondents), the study sample was 2866 older adults (1433 with and 1433 without VI), representing 2.67 million older adults in the United States.

Sample Characteristics

As Table 1 shows, of the study sample, 39.94%, 34.37%, and 25.68% of the VI group were in the 65-74, 75-84, and ≥85 age groups, respectively. The proportion of those aged ≥85 years in the VI group was much higher than the proportion of the same age group among all NHIS sample adults (11.38%). As the VI and no VI groups were matched for age, sex, and number of medical conditions, these variables were not significantly different between the 2 groups. However, even after matching, compared with the no VI group, the VI group had significantly higher proportions of racial/ethnic minorities, nonmarried individuals, and those with an income less than 200% of the poverty threshold and Medicaid and lower proportions of those with a college degree and private health insurance.

Of the VI group, 88.15% reported that they had VI for more than a year, and 96.99% reported that VI was a chronic condition. The VI group also included more blind and hearing-impaired individuals, with 31.01% (compared with 2.87% in the no VI group) reporting hearing-caused difficulty with activities. Psychological distress was significantly higher among the VI group, with approximately one-fourth of participants reporting mild/moderate or serious psychological distress in the preceding month. The VI and no VI groups did not differ in rates of hospitalization and visits/consultations with health care professionals other than eye doctors. Although the data did not show the exact number of times these older adults saw/talked with health care providers, a higher proportion of the VI group reported receiving health care services ≥10 times, indicating that they were more frequent health care service users than those without VI.

Table 1. Sociodemographic, health insurance, and health care use characteristics of the matched sample

Variables	No VI[a] (n=1433)	VI (n=1433)	*P* value[b]
Age (years), %			.67
65-74	40.34	39.94	
75-84	32.51	34.37	
85+	27.15	25.68	
Female, %	59.21	60.43	.62
Race/ethnicity, %			<.001
Non-Hispanic white	78.75	68.79	
Non-Hispanic black	10.04	11.09	
Hispanic	6.98	10.88	
Non-Hispanic Asian	3.09	7.04	
Other	1.14	2.20	
Married, %	49.12	42.96	.02
College degree, %	23.36	18.72	.02
Family income: percentage of poverty			<.001
<200	32.59	43.66	
200-399	31.99	25.08	
≥400	26.66	21.97	
Missing	8.76	9.29	
VI duration, %			N/A[c]
>3 months	N/A	1.83	
3-5 months	N/A	0.80	
6-12 months	N/A	8.04	
More than a year	N/A	88.15	
Missing	N/A	1.18	
VI as a chronic problem, %			N/A
Yes	N/A	96.99	
No	N/A	3.01	
Blind, %	0.38	12.30	<.001
Hearing impairment, %	2.87	31.01	<.001
Number of medical conditions, mean (SE)	3.14 (0.06)	3.12 (0.06)	.77
Psychological distress (K6[d] score)[e]			<.001
Mean (SE)	2.70 (0.14)	4.53 (0.20)	
No distress (K6 score=0-7), %	88.21	75.94	
Mild/moderate distress (K6 score=8-12), %	7.96	14.14	
Serious distress (K6 score=13-24), %	3.83	9.92	
Health insurance, %			
Medicare	95.48	93.72	.17
Medicaid	11.08	21.58	<.001
Private health insurance	44.40	34.70	<.001
Military health insurance	8.86	10.43	.26
Health care use in the past 12 months, %			

Variables	No VI[a] (n=1433)	VI (n=1433)	*P* value[b]
Hospitalized	23.08	23.49	.83
Received care 10 or more times	27.97	33.47	.02
Saw/talked to a general practitioner	89.38	87.77	.30
Saw/talked to an eye doctor	63.09	71.99	<.001
Saw/talked to a medical specialist	52.58	47.51	.05
Saw/talked to a physical/occupational therapist	20.36	23.25	.16
Saw/talked to a mental health professional	5.97	7.88	.14

[a] VI: vision impairment.

[b] Probability values for differences between the no VI and VI groups were calculated using Pearson chi-square tests for categorical variables and, two-tailed, independent sample *t* tests for continuous variables (number of medical conditions and K6 scores).

[c] N/A: not applicable.

[d] K-6: Kessler-6 screen.

[e] Sample size is 2809 (1413 without VI and 1396 with VI) because of missing data.

Internet and Health Information Technology Use

Table 2 shows that compared with 42.61% and 30.87% of the no VI group, 29.63% and 22.37% of the VI group used the internet and any HIT, respectively. More specifically, the VI group was significantly less likely to have sought health information on the web or communicated with health care providers via email; however, the 2 groups did not differ with respect to filling a prescription and scheduling a medical appointment on the web. In addition to health information seeking, only a small proportion (3.69%-7.96%) of these older adults with or without VI used HIT for other medical/health purposes. Table 2 also shows that of those who used the internet, the VI group and the no VI group did not differ on the rates of HIT use: 67.04% of the no VI group and 69.58% of the VI group used any HIT. Although internet users' most common HIT use was for health information seeking, between 10.80% and 18.31% also engaged in the other 3 types of HIT use.

Table 2. Internet and health information technology use in the past 12 months among the matched sample

Variable	No VI[a] (n=1433), % (95% CI)	VI (n=1433), % (95% CI)	*P* value[b]
Internet use	42.61 (39.37-45.93)	29.63 (26.50-32.96)	<.001
HIT[c] use	30.87 (27.95-33.94)	22.37 (19.53-25.48)	<.001
Looked up health information on the internet	28.59 (25.74-31.63)	20.54 (17.83-23.54)	<.001
Filled a prescription on the internet	6.71 (5.24-8.55)	4.91 (3.67-6.55)	.11
Scheduled medical appointment on the internet	5.66 (4.18-7.62)	3.69 (2.52-5.38)	.07
Communicated with health care provider by email	7.96 (6.21-10.13)	5. 45 (4.14-7.13)	.04
HIT use among internet users[d]	67.04 (62.32-71.43)	69.58 (63.42-75.11)	.50
Looked up health information on the internet	61.81 (56.82-66.56)	65.62 (59.44-71.32)	.33
Filled a prescription on the internet	15.72 (12.37-19.78)	15.54 (11.60-20.51)	.95
Scheduled medical appointment on the internet	12.86 (9.56-17.08)	10.80 (7.38-15.56)	.46
Communicated with health care provider by email	18.31 (14.47-22.89)	16.42 (12.43-21.39)	.55

[a] VI: vision impairment.

[b] Probability values for differences between the no VI and VI groups were calculated using Pearson chi-square tests.

[c] HIT: health information technology.

[d] Sample size is 963 (582 without VI and 381 with VI).

Association of Vision Impairment With Internet and Health Information Technology Use Multivariable Analyses

Table 3 shows that controlling for sociodemographic characteristics, health status, health insurance, and health care service use, VI was still significantly associated with lower odds of internet (aOR 0.64, 95% CI 0.49-0.83) and HIT (aOR 0.74, 95% CI 0.56-0.97) use, although being blind by itself was not a significant factor. In terms of other significant covariates, being in the 2 older age groups, non-Hispanic black or Hispanic, and nonmarried and receiving health care ≥10 times were associated with lower odds of both internet and HIT use, whereas having a college degree, family income ≥200% of the poverty threshold (compared with <200% of the poverty threshold), and veterans/military insurance and seen/talked to a medical specialist and a PT/OT were associated with greater odds of both internet and HIT use. In addition, receiving Medicaid and having seen/talked to a mental health provider were associated with lower odds, and having seen/talked to a general practitioner and an eye doctor was associated with higher odds of internet use but not HIT use.

Table 3 also shows that among internet users, VI was not a significant factor for HIT use. Age group, marital status, education, and health insurance type were not significant. Compared with non-Hispanic white internet users, only Hispanic internet users had lower odds of HIT use. With respect to family income, only those in the *missing* income category had significantly greater odds than those with income <200% of poverty to have used any HIT. Having seen/talked to a medical specialist and a PT/OT was also associated with greater odds of HIT use among internet users.

Table 3. Association of vision impairment with internet use and any health information technology use: logistic regression results.

Variable[a]	Internet use vs nonuse, aOR[b] (95% CI)	Any HIT[c] use vs nonuse, aOR[b] (95% CI)	Any HIT use vs nonuse among internet users, aOR[b] (95% CI)
Vision impairment vs no impairment	0.64 (0.49-0.83)[d]	0.74 (0.56-0.97)[e]	1.15 (0.77-1.74)
Blindness vs no blindness	0.83 (0.44-1.57)	1.08 (0.54-2.17)	2.32 (0.72-7.43)
Age group (years; vs 65-74 years)			
75-84	0.39 (0.29-0.52)[f]	0.45 (0.34-0.59)[f]	0.92 (0.61-1.37)
≥85	0.20 (0.15-0.27)[f]	0.17 (0.11-0.25)[f]	0.36 (0.20-0.63)
Female vs male	1.31 (1.00-1.71)	1.25 (0.96-1.64)	1.02 (0.69-1.49
Race/ethnicity (vs non-Hispanic white)			
Non-Hispanic black	0.35 (0.23-0.54)[f]	0.44 (0.28-0.71)[d]	0.87 (0.40-1.86)
Hispanic	0.23 (0.14-0.40)[f]	0.23 (0.12-0.45)[f]	0.37 (0.16-0.83)[e]
Non-Hispanic Asian	0.53 (0.31-0.91)[e]	0.86 (0.41-1.77)	1.81 (0.50-6.54)
Other	1.07 (0.53-2.15)	0.89 (0.41-1.94)	0.69 (0.22-2.19)
Not married vs married	0.70 (0.55-0.89)[d]	0.76 (0.59-0.99)[e]	1.00 (0.70-1.43)
College degree vs no degree	4.17 (3.09-5.62)[f]	3.04 (2.23-4.12)[f]	1.27 (0.86-1.87)
Family income (vs <200% of the poverty threshold)			
200-399	1.37 (1.03-1.83)[e]	1.53 (1.11-2.12)[f]	1.23 (0.75-2.04)
≥400	1.93 (1.40-2.66)[f]	2.01 (1.41-2.86)[g]	1.46 (0.88-2.43)
Missing	1.01 (0.67-1.53)	1.41 (0.90-2.22)	2.71 (1.11-6.62)[f]
Number of medical conditions	0.96 (0.88-1.03)	1.00 (0.92-1.08)	1.08 (0.96-1.22)
Hearing impairment vs no impairment	1.09 (0.75-1.59)	0.97 (0.66-1.43)	0.95 (0.54-1.69)
Medicaid vs no Medicaid	0.63 (0.41-0.96)[e]	0.76 (0.48-1.19)	1.24 (0.63-2.45)
Private HI[g] vs no private HI	1.07 (0.83-1.39)	1.15 (0.88-1.52)	1.25 (0.63-2.45)
Military HI vs no military HI	1.58 (1.08-2.31)[e]	1.53 (1.02-2.31)[e]	0.87 (0.47-1.60)
Hospitalized vs not hospitalized	0.92 (0.68-1.25)	0.87 (0.61-1.24)	0.80 (0.51-1.24)
Received care ≥10 times vs received no/less care	0.71 (0.54-0.92)[e]	0.88 (0.65-1.18)	1.22 (0.79-1.88)
Saw/talked to a general practitioner vs did not see/talk	1.80 (1.21-2.67)[d]	1.27 (0.84-1.93)	0.90 (0.35-2.04)
Saw/talked to an eye doctor vs did not see/talk	1.35 (1.04-1.74)[e]	1.31 (0.99-1.74)	1.01 (0.68-1.52)
Saw/talked to a medical specialist vs did not see/talk	1.50 (1.16-1.95)[d]	1.81(1.40-2.35)[f]	1.80 (1.21-2.68)[d]
Saw/talked to a physical/occupational therapist vs did not see/talk	1.41 (1.04-1.91)[e]	1.51 (1.11-2.05)[e]	1.57 (1.00-2.44)[e]
Saw/talked to a mental health professional vs did not see/talk	0.53 (0.33-0.85)[d]	0.86 (0.57-1.30)	1.48 (0.73-3.00)

[a] Model statistics: N=2866, design *df*=885, F (26, 860)=14.54, *P* value <.001 for internet use versus nonuse; N=2866, design *df*=885, F (26, 860)=10.90, *P* value <.001 for any HIT use vs nonuse; and N=963, design *df*=794, F (26, 860)=2.13; *P* value <.001 for any HIT use vs nonuse among internet users.

[b] aOR: adjusted odds ratio.

[c] HIT: health information technology.

[d] *P*<.01.

[e] *P*<.05.

[f] *P*<.001.

[g] HI: health insurance.

Association of Health Information Technology Use With Psychological Distress: Multivariable Analyses

Table 4 shows that controlling for sociodemographic characteristics, health status, health insurance, and health care service use, HIT use was significantly associated with lower odds of mild/moderate or serious psychological distress (aOR 0.62, 95% CI 0.43-0.90). As expected, VI was associated with greater odds of mild/moderate or serious psychological distress (aOR 1.84, 95% CI 1.36-2.49), but blindness by itself was not. Of the other covariates, being in the 2 older age groups and black, and having a college degree and private health insurance were associated with lower odds of mild/moderate or serious distress; however, being female and nonmarried, having a higher number of medical conditions and hearing impairment, and being seen/having talked to a mental health provider were associated with greater odds of mild/moderate or serious distress. The interaction terms between HIT use and VI were not significant. The findings from the logistic regression model with internet use as the dependent variable were similar, showing lower odds of mild/moderate or serious distress among internet users but higher odds of such distress among those with VI.

Table 4. Association of any health information technology use with psychological distress: logistic regression results.

Variable[a]	Mild/moderate or serious vs no distress, aOR[b] (95% CI)
Any health information technology use	0.62 (0.43-0.90)[c]
Vision impairment vs no impairment	1.84 (1.36-2.49)[d]
Blindness vs no blindness	1.44 (0.83-2.51)
Age group (vs 65-74 years)	
75-84	0.53 (0.38-0.74)[d]
≥85	0.46 (0.31-0.70)[d]
Female vs male	1.40 (1.02-1.91)[c]
Race/ethnicity (vs non-Hispanic white)	
Non-Hispanic black	0.61 (0.39-0.93)[c]
Hispanic	1.10 (0.68-1.76)
Non-Hispanic Asian	1.18 (0.61-2.28)
Other	0.76 (0.37-1.56)
Not married vs married	1.40 (1.04-190)[c]
College degree vs no degree	0.60 (0.39-0.92)[c]
Family income (vs <200% of the poverty threshold)	
200-399	1.09 (0.76-1.55)
≥400+	0.95 (0.65-1.38)
Missing	0.74 (0.43-1.28)
Number of medical conditions	1.15 (1.05-1.26)[e]
Hearing impairment vs no impairment	1.90 (1.38-2.61)[d]
Medicaid vs no Medicaid	1.18 (0.79-1.75)
Private HI[f] vs no private HI	0.87 (0.63-1.20)
Military HI vs no military HI	0.68 (0.41-1.15)
Hospitalized vs not hospitalized	1.10 (0.78-1.56)
Received care ≥10 times vs received no/less care	1.30 (0.97-1.74)
Saw/talked to a general doctor vs did not see/talk	1.40 (0.89-2.20)
Saw/talked to an eye doctor vs did not see/talk	0.80 (0.59-1.07)
Saw/talked to a medical specialist vs did not see/talk	0.91 (0.68-1.22)
Saw/talked to a physical/occupational therapist vs did not see/talk	1.01 (0.73-1.39)
Saw/talked to a mental health professional vs did not see/talk	3.83 (2.44-5.99)[d]

[a] Model statistics: N=2809, design *df*=885, F (27, 859)=6.40, *P* value <.001.

[b] aOR: adjusted odds ratio.

[c] *P*<.05.

[d] *P*<.001.

[e] *P*<.01.

[f] HI: health insurance.

Discussion

Principal Findings

This study shows that 3.28% of older adults in the United States (compared with 0.84% of those aged 18-64 years) reported that their VI caused difficulty with activities. As expected, the ≥85 age group was overrepresented among those with VI as common age-related eye disorders (eg, macular degeneration, glaucoma, and cataract) tend to be associated with a gradual decline in vision [26]. Although these older adults with VI were matched in a 1:1 ratio with their peers without VI based on age, gender, and health/other disability status, those with VI were still significantly more socioeconomically disadvantaged in terms of education and income, included more racial/ethnic minorities, and had higher proportions of those with blindness and hearing impairment. They also used more health care in the past year.

The first key finding is that older adults with VI are less likely to engage in internet and HIT use than their matched age peers, even after controlling for other sociodemographic, health status, health insurance, and health care use variables, which supports H1. However, among those who used the internet, VI was not associated with the odds of HIT use. This indicates that among older adults with VI, their lower likelihood of internet use is the main reason for their overall lower rate of HIT use compared with their matched age peers. Internet users with VI are likely to navigate the internet using screen magnification and reading software and other accessibility features for vision [27] and are as likely to use HIT as their peers without VI. The second key finding is that HIT use is associated with a lower likelihood of having experienced mild/moderate or serious psychological distress in the past month, controlling for VI, which supports H2. VI is associated with greater odds of having experienced mild/moderate or severe psychological distress, suggesting that VI (and hearing impairment) negatively impacts overall well-being.

The lower odds of internet use among older adults with VI may be because of their lower socioeconomic status and higher levels of health problems and disability that previous studies identified as major contributors to the digital divide [7,8]. Costs of an internet subscription and equipment (eg, computer, tablet) are not likely to be priority items, especially among older adults with a low income who have to contend with the financial strains of managing their health and disability (eg, costs of prescription medications and transportation to health care appointments). Given that almost one-third of older adults with VI also had hearing impairment, these double disabilities were likely barriers to internet and HIT use. A survey of individuals aged 50 to 74 years in the United Kingdom found that those with moderate/serious hearing difficulties were less likely to use computers than those with no hearing difficulty [28].

Our findings also show that internet use was higher if older adults saw/talked to a general practitioner, eye doctor, medical specialist, or PT/OT, and HIT use was higher if they saw/talked to a medical specialist or PT/OT. Those who had visits with an eye doctor may be more socioeconomically advantaged than those who did not. Ehrlich et al [29] found that unmarried, older adults with a low income were less likely than their more advantaged peers to report a recent eye examination, and common reasons for not having an eye examination included cost and lack of insurance coverage. Thus, the relationship between health care provider contact and internet use may reflect the digital divide because of socioeconomic status. The association of HIT use with specialist and PT/OT consultations is also consistent with a previous study [7], suggesting that patients with complex medical conditions that lead to specialist care may be more likely to use HIT to better understand their conditions and treatments. Physical and occupational therapists also tend to spend more time with patients than physician providers do, which allows them more opportunities for patient education. Thus, patient education may have extended the use of web-based resources and encouragement for older adult patients to use these resources as part of promoting specific condition- and treatment-related knowledge as well as overall health to prevent disease, disability, and injury. The importance of integrating health promotion and wellness in PT/OT practice has also been underscored [30].

The finding that internet/HIT use was higher if older adults had veterans/military health insurance regardless of VI is likely because of the fact that the Veterans Administration (VA) health care systems are the most digitalized and telehealth-oriented of all health care systems [31]. VA patients may receive greater encouragement to use HIT to seek health information and navigate the VA health care systems. More research is needed on specific health care service/system use and HIT use among older adults.

Along with the costs of technology adoption and continued connectivity, lack of interest in and distrust of the internet as a source of health information and means of health management may be barriers to the use of HIT by older adults. A survey done in the United Kingdom on the use/nonuse of consumer electronic devices (eg, smartphones, tablet computers, and electronic book [ebook] readers) among people with VI found that cost and lack of interest were among the most frequently cited reasons for nonuse [32]. An earlier study also found that older adults were less likely than middle-aged adults to trust the internet as a source of health information; however, the association between age and distrust was no longer significant after adjusting for potential contributors to distrust, such as confusion in using the internet and providing too much information [33]. Lack of interest may also be caused by a lack of knowledge and skills for adopting technology.

Given the divergent findings discussed earlier on the relationship between ICT use and psychological well-being among older adults, more research is needed to examine moderators and mediators of the association between HIT use and lower likelihood of mild/moderate or serious psychological distress found in the study sample. Among those with VI, ICT use may facilitate social connections (eg, participation in online support and interest groups) and informational/instrumental activities (eg, reading/listening to ebooks, watching/listening to live-streamed religious services). Studies show that ICT use among older adults is associated with increased engagement/participation in social, instrumental, and leisure activities, with some ICT use compensating for aging-related and other challenging circumstances [34,35].

Limitations

This study has limitations because of data constraints. First, VI was self-reported, and data on age of onset and change over time were not available. Thus, we do not know if some older adults with VI had stopped using the internet because of progressing VI. Second, because the NHIS reports HIT use as a dichotomous variable (use vs nonuse), we could not study the effects of HIT use frequency. More detailed data on HIT use frequency and context can help in better understanding their effects on the well-being of older adults. Third, additional data on the type of activities negatively affected by VI (and other disabilities) would also have been helpful in better understanding their impacts and recommending strategies that can help people with disabilities use the internet and HIT to their advantage. Fourth, because the NHIS data are cross-sectional, we can report associations (correlations) but not causal relationships.

Conclusions

As more health care services and instrumental daily living activities (eg, banking, paying bills, applying for social services) require ICT use, the digital divide among older adults with VI must be closed. Given the potential of ICT to help older adults with VI remain active and enrich their lives, closing the digital divide may also improve their psychological well-being. First, we recommend that socioeconomically disadvantaged older adults experiencing the digital divide should be provided help to access ICT through a fee waiver or subsidy to cover internet equipment (eg, computer, tablet) and subscription services. The movement for equitable internet access as a public utility has gained momentum in state and local governments [36]. Older adults should be included as targets as they are the largest group of consumers of health care services and services that increasingly rely on ICT use. Second, older adults with VI who do not know how to use the internet and HIT but want to learn should be provided instruction and technology support services. Attention must be paid to accessibility features and adaptive devices that will facilitate technology use among those with VI. Third, in addition to improved ICT access, older adults with a low income need better access to preventive eye care and treatment of VI and other health care services. As refractive errors and some age-related eye diseases can be effectively treated to improve vision, more concerted efforts are needed to reduce VI in late life. Improved access to other health care services and providers is also important for patient education, prevention, treatment, and alleviation of health problems and functional impairment. As our study shows, certain health care provider contacts are associated with higher HIT use among older adults, suggesting attempts to better understand their health problems and treatments using technology.

Conflicts of Interest

None declared.

References

1. American Foundation for the Blind. 2016. Statistics about Seniors with Vision Loss URL: https://www.afb.org/research-and-initiatives/statistics/statistics-seniors-vision-loss [accessed 2019-11-01]
2. Okoro CA, Hollis ND, Cyrus AC, Griffin-Blake S. Prevalence of disabilities and health care access by disability status and type among adults-United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Aug 17;67(32):882-887 [FREE Full text] [doi: 10.15585/mmwr.mm6732a3] [Medline: 30114005]
3. Swenor BK, Lee MJ, Varadaraj V, Whitson HE, Ramulu PY. Aging with vision loss: a framework for assessing the impact of visual impairment on older adults. Gerontologist 2019 Aug 24:- epub ahead of print. [doi: 10.1093/geront/gnz117] [Medline: 31504483]
4. Pelletier AL, Rojas-Roldan L, Coffin J. Vision loss in older adults. Am Fam Physician 2016 Aug 1;94(3):219-226 [FREE Full text] [Medline: 27479624]
5. Anderson M, Perrin A. Pew Research Center. 2017. Technology Use Among Seniors URL: https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/ [accessed 2019-11-19]
6. Hong YA, Cho J. Has the digital health divide widened? Trends of health-related internet use among older adults from 2003 to 2011. J Gerontol B Psychol Sci Soc Sci 2017 Sep 1;72(5):856-863. [doi: 10.1093/geronb/gbw100] [Medline: 27558403]
7. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US national health interview survey. J Med Internet Res 2011 Apr 20;13(2):e33 [FREE Full text] [doi: 10.2196/jmir.1753] [Medline: 21752784]
8. Hall AK, Bernhardt JM, Dodd V, Vollrath MW. The digital health divide: evaluating online health information access and use among older adults. Health Educ Behav 2015 Apr;42(2):202-209 [FREE Full text] [doi: 10.1177/1090198114547815] [Medline: 25156311]
9. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, ehealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May 2;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
10. Keränen NS, Kangas M, Immonen M, Similä H, Enwald H, Korpelainen R, et al. Use of information and communication technologies among older people with and without frailty: a population-based survey. J Med Internet Res 2017 Feb 14;19(2):e29 [FREE Full text] [doi: 10.2196/jmir.5507] [Medline: 28196791]
11. Burns P, Jones SC, Caputi P, Iverson D. Are older Australians with chronic diseases online? Health Promot J Austr 2018 Apr;29(1):72-78. [doi: 10.1002/hpja.5] [Medline: 29700931]
12. Shim H, Ailshire J, Zelinski E, Crimmins E. The health and retirement study: analysis of associations between use of the internet for health information and use of health services at multiple time points. J Med Internet Res 2018 May 25;20(5):e200 [FREE Full text] [doi: 10.2196/jmir.8203] [Medline: 29802088]

13. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):763-771. [doi: 10.1093/geronb/gbu018] [Medline: 24671896]
14. Forsman AK, Nordmyr J. Psychosocial links between internet use and mental health in later life: a systematic review of quantitative and qualitative evidence. J Appl Gerontol 2017 Dec;36(12):1471-1518. [doi: 10.1177/0733464815595509] [Medline: 26245208]
15. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015 May;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
16. Sims T, Reed AE, Carr DC. Information and communication technology use is related to higher well-being among the oldest-old. J Gerontol B Psychol Sci Soc Sci 2017 Sep 1;72(5):761-770. [doi: 10.1093/geronb/gbw130] [Medline: 27702839]
17. Fang Y, Chau AK, Wong A, Fung HH, Woo J. Information and communicative technology use enhances psychological well-being of older adults: the roles of age, social connectedness, and frailty status. Aging Ment Health 2018 Nov;22(11):1516-1524. [doi: 10.1080/13607863.2017.1358354] [Medline: 28777010]
18. Fang Y, Chau AK, Fung HH, Woo J. Loneliness shapes the relationship between information and communications technology use and psychological adjustment among older adults. Gerontology 2019;65(2):198-206. [doi: 10.1159/000495461] [Medline: 30814472]
19. Khalaila R, Vitman-Schorr A. Internet use, social networks, loneliness, and quality of life among adults aged 50 and older: mediating and moderating effects. Qual Life Res 2018 Feb;27(2):479-489. [doi: 10.1007/s11136-017-1749-4] [Medline: 29210015]
20. Elliot AJ, Mooney CJ, Douthit KZ, Lynch MF. Predictors of older adults' technology use and its relationship to depressive symptoms and well-being. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):667-677 [FREE Full text] [doi: 10.1093/geronb/gbt109] [Medline: 24304556]
21. Centers for Disease Control and Prevention. 2019. NHIS Data, Questionnaires and Related Documentation URL: https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm [accessed 2019-10-10]
22. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003 Feb;60(2):184-189. [doi: 10.1001/archpsyc.60.2.184] [Medline: 12578436]
23. Kessler RC, Galea S, Gruber MJ, Sampson NA, Ursano RJ, Wessely S. Trends in mental illness and suicidality after hurricane Katrina. Mol Psychiatry 2008 Apr;13(4):374-384 [FREE Full text] [doi: 10.1038/sj.mp.4002119] [Medline: 18180768]
24. Cook D. Web Log. 2015 Jan 28. Ccmatch URL: https://github.com/danielecook/ccmatch [accessed 2018-04-20]
25. Allison P. Statistical Horizons. 2012 Sep 10. When Can You Safely Ignore Multicollinearity? URL: https://statisticalhorizons.com/multicollinearity [accessed 2020-04-21]
26. Chen SP, Bhattacharya J, Pershing S. Association of vision loss with cognition in older adults. JAMA Ophthalmol 2017 Sep 1;135(9):963-970 [FREE Full text] [doi: 10.1001/jamaophthalmol.2017.2838] [Medline: 28817745]
27. American Foundation for the Blind. The Visually Impaired Web User's Technology URL: https://www.afb.org/about-afb/what-we-do/afb-consulting/afb-accessibility-resources/users-technology [accessed 2019-11-15]
28. Henshaw H, Clark DP, Kang S, Ferguson MA. Computer skills and internet use in adults aged 50-74 years: influence of hearing difficulties. J Med Internet Res 2012 Aug 24;14(4):e113 [FREE Full text] [doi: 10.2196/jmir.2036] [Medline: 22954484]
29. Ehrlich JR, Ndukwe T, Solway E, Woodward MA, Singer DC, Newman-Casey PA, et al. Self-reported eye care use among us adults aged 50 to 80 years. JAMA Ophthalmol 2019 Jun 20:- epub ahead of print. [doi: 10.1001/jamaophthalmol.2019.1927] [Medline: 31219510]
30. Morris D, Jenkins G. Preparing physical and occupational therapists to be health promotion practitioners: a call for action. Int J Environ Res Public Health 2018 Feb 24;15(2):392 [FREE Full text] [doi: 10.3390/ijerph15020392] [Medline: 29495307]
31. Eddy N. Healthcare IT News. 2019 Nov 25. VA Sees a Surge in Veterans' Use of Telehealth Services URL: https://www.healthcareitnews.com/news/va-sees-surge-veterans-use-telehealth-services [accessed 2019-12-01]
32. Crossland MD, Silva RS, Macedo AF. Smartphone, tablet computer and e-reader use by people with vision impairment. Ophthalmic Physiol Opt 2014 Sep;34(5):552-557. [doi: 10.1111/opo.12136] [Medline: 25070703]
33. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011 Feb 16;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
34. Kim J, Lee HY, Christensen MC, Merighi JR. Technology access and use, and their associations with social engagement among older adults: do women and men differ? J Gerontol B Psychol Sci Soc Sci 2017 Sep 1;72(5):836-845. [doi: 10.1093/geronb/gbw123] [Medline: 28073816]
35. Nimrod G. Aging well in the digital age: technology in processes of selective optimization with compensation. J Gerontol B Psychol Sci Soc Sci 2019 Aug 27:- epub ahead of print. [doi: 10.1093/geronb/gbz111] [Medline: 31504873]
36. Berman D, Pickard V. The Conversation. 2019 Nov 14. Cities and States Take Up the Battle for an Open Internet URL: https://theconversation.com/cities-and-states-take-up-the-battle-for-an-open-internet-126170 [accessed 2019-11-20]

Abbreviations

aOR: adjusted odds ratio
BRFSS: Behavioral Risk Factor Surveillance System
CDC: Centers for Disease Control and Prevention
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
ebook: electronic book
HIT: health information technology
ICT: information and communications technology
K6: Kessler-6 screen
NHIS: National Health Interview Survey
PT/OT: physical therapists/occupational therapists
VA: Veterans Administration
VI: vision impairment

Edited by G Eysenbach; submitted 03.12.19; peer-reviewed by H Krampe, R Yin, H Similä; comments to author 30.01.20; revised version received 02.02.20; accepted 29.03.20; published 03.06.20.

Please cite as:
Choi NG, DiNitto DM, Lee OE, Choi BY
Internet and Health Information Technology Use and Psychological Distress Among Older Adults With Self-Reported Vision Impairment: Case-Control Study
J Med Internet Res 2020;22(6):e17294
URL: https://www.jmir.org/2020/6/e17294
doi:10.2196/17294
PMID:32490851

Review

Health Equity in the Effectiveness of Web-Based Health Interventions for the Self-Care of People With Chronic Health Conditions: Systematic Review

Sophie Turnbull[1], BSc, MSc, PhD; Christie Cabral[1], BSc, MRES, PhD; Alastair Hay[1], MBChB, MD, MRCP, DCH, MRCGP, FRCGP, FRCPCH (Hons); Patricia J Lucas[2], BA, PhD

[1]Centre for Academic Primary Care, Population Health Sciences, University of Bristol, Bristol, United Kingdom
[2]School for Policy Studies, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Sophie Turnbull, BSc, MSc, PhD
Centre for Academic Primary Care
Population Health Sciences
University of Bristol
Canynge Hall
39 Whatley Road
Bristol,
United Kingdom
Phone: 44 117 331 0590
Email: sophie.turnbull@bristol.ac.uk

Abstract

Background: Web-based self-care interventions have the potential to reduce health inequalities by removing barriers to access to health care. However, there is a lack of evidence about the equalizing effects of these interventions on chronic conditions.

Objective: This study investigated the differences in the effectiveness of web-based behavioral change interventions for the self-care of high burden chronic health conditions (eg, asthma, chronic obstructive pulmonary disease [COPD], diabetes, and osteoarthritis) across socioeconomic and cultural groups.

Methods: A systematic review was conducted, following Cochrane review guidelines. We conducted searches in Ovid Medical Literature Analysis and Retrieval System Online and Cumulative Index to Nursing and Allied Health Literature databases. Studies with any quantitative design were included (published between January 1, 2006, and February 20, 2019) if they investigated web-based self-care interventions targeting asthma, COPD, diabetes, and osteoarthritis; were conducted in any high-income country; and reported variations in health, behavior, or psychosocial outcomes across social groups. Study outcomes were investigated for heterogeneity, and the possibility of a meta-analysis was explored. A narrative synthesis was provided together with a novel figure that was developed for this review, displaying heterogeneous outcomes.

Results: Overall, 7346 records were screened and 18 studies were included, most of which had a high or critical risk of bias. Important study features and essential data were often not reported. The meta-analysis was not possible due to the heterogeneity of outcomes. There was evidence that intervention effectiveness was modified by participants' social characteristics. Minority ethnic groups were found to benefit more from interventions than majority ethnic groups. Single studies with variable quality showed that those with higher education, who were employed, and adolescents with divorced parents benefited more from interventions. The evidence for differences by age, gender, and health literacy was conflicting (eg, in some instances, older people benefited more, and in others, younger people benefited more). There was no evidence of differences in income, numeracy, or household size.

Conclusions: There was evidence that web-based self-care interventions for chronic conditions can be advantageous for some social groups (ie, minority ethnic groups, adolescents with divorced parents) and disadvantageous for other (ie, low education, unemployed) social groups who have historically experienced health inequity. However, these findings should be treated with caution as most of the evidence came from a small number of low-quality studies. The findings for gender and health literacy were mixed across studies on diabetes, and the findings for age were mixed across studies on asthma, COPD, and diabetes. There was no evidence that income, numeracy, or the number of people living in the household modified intervention effectiveness.

We conclude that there appear to be interaction effects, which warrant exploration in future research, and recommend a priori consideration of the predicted interaction effects.

Trial Registration: PROSPERO CRD42017056163; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=56163

(J Med Internet Res 2020;22(6):e17849) doi:10.2196/17849

KEYWORDS

health equity; self-care; eHealth; intervention; diabetes; asthma; pulmonary disease, chronic obstructive; osteoarthritis

Introduction

Chronic Illness and Health Inequalities

Chronic or long-term conditions such as diabetes have a significant impact on the individual's quality of life and are the major cause of disability and premature death worldwide [1,2]. In high-income countries, chronic conditions are estimated to account for 87% of all deaths [3]. Due to their nature, they cause illness over long periods and their management is complex and costly [2]. In countries where health care is universally provided (such as the National Health Service [NHS] in the United Kingdom), services are struggling with the increasing demand, partly driven by the increasing number of people with chronic conditions [4-7]. Even in high-income countries, people living in constrained conditions and with a lower socioeconomic status (SES) experience chronic illness more commonly and severity is greater than average [2,8]. For example, 52% of those from unskilled occupations suffer from chronic conditions in comparison with 33% of those in professional occupations [9]. These inequalities in health have been attributed to social determinants of health (SDH) and inequity in access to health care [10]. SDH are the complex interacting elements in the physical and social environment that contribute toward disparities in health status. Inequalities in the distribution of good quality health care mean that people do not have equal access to treatments that can improve health outcomes. Taken together, SDH mean disadvantaged groups suffer more illness and more severe illness, but are least likely to receive effective treatment, which together result in disparities in health outcomes [11].

Proposed Solutions to Increase Access to Health Care for People With Chronic Conditions

Both self-care and web-based interventions have been proposed as methods for increasing access to health care while relieving pressure on health care services. The underlying assumption with self-care interventions is that they provide health care where there was none, by encouraging people to be their own health resource [12]. Web-based interventions have the potential to increase access to good quality health care by providing support to an almost unlimited number of people from the same digital platform at the same time, and the interventions can be tailored to individual needs [13-16]. More recently, combinations of the 2 approaches, in the form of web-based self-care interventions, have become more prevalent for a range of health conditions, particularly for chronic conditions [17,18]. These digital self-care interventions are viewed as playing a vital role in the prevention and treatment of long-term illnesses such as chronic lung disease, cancer, cardiovascular disease, and diabetes [18,19].

The Impact of Electronic Health and Self-Care Interventions on Health Inequalities

Evidence of the impact of these approaches on health inequalities has been mixed. Nondigital self-care interventions in the form of community-based training courses have been found to improve health status, health behaviors, and the quality of life of patients with chronic conditions [20-24]. However, there is evidence that disadvantaged groups face barriers to accessing these interventions, such as the high levels of health literacy that are often required to understand the training materials as well as language barriers where the training is only conducted in English [8]. Web-based interventions designed specifically for those from underserved and disadvantaged groups have been found to benefit these populations [14,25,26]. However, there is also evidence that access and usability for disadvantaged groups remain to be barriers [27-30]. People from lower socioeconomic groups and older adults are less likely to seek health information on the web and have problems using the web-based information available [31-38]. There is an absence of systematic review evidence investigating whether web-based self-care interventions designed for people with chronic conditions are equally effective for people with different social characteristics.

Methods

A systematic review was conducted following the Cochrane review guidelines [39] and was reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses-Equity 2012 extension checklist [40]. The protocol was registered in advance on the PROSPERO international prospective register of systematic reviews (registration number CRD42017056163).

Objective

This study aimed to synthesize evidence investigating whether web-based self-care interventions designed for people with chronic conditions are equally effective for people with different social characteristics.

Inclusion Criteria

Population

A total of 4 high burden physical health conditions were included: asthma, diabetes mellitus (type 1 and type 2), osteoarthritis, and chronic obstructive pulmonary disease (COPD). The health conditions were identified using the World Health Organization's disease burden data and were selected from the top 10 conditions that cause the greatest number of years lost to disease [2,41-44]. All included conditions cause considerable burden and disability to patients [41] and health services [42,43] and have been shown to have social patterning in severity and incidence [2,44]. Furthermore, all 4 have the potential for symptoms, severity, and prognosis to be improved through changes to behavior, such as diet or physical activity.

Intervention

We included interventions that were aimed at improving symptoms or prognosis and had a web-based component or were delivered exclusively on the web. This included mobile apps with web connectivity. Interventions were included if they were predominantly reliant on the individual changing their self-care behavior without intensive contact with a therapist or clinician.

Study Types

Studies with a quantitative design were included, such as randomized controlled trials (RCTs), observational studies, quasi-experimental designs, feasibility and pilot studies, and mixed methods studies that included a quantitative element. Abstracts were not included where no full publication or report was available [45].

Systematic Review Outcomes

Studies reporting health, behavior, knowledge, and psychosocial outcomes were included.

Available Data

Studies were included if the study teams had explored whether the social or cultural groups had modified intervention effectiveness and whether the independent contribution of the group on the outcome could be determined. The authors were contacted for models where the independent contribution of the social group could not be determined in the text.

Exclusion Criteria

There were no language restrictions. The publication dates were limited from January 1, 2006, to February 20, 2019, to ensure that the review included interventions with recent technology.

Search Strategy

We conducted searches in Ovid Medical Literature Analysis and Retrieval System Online (MEDLINE; Allied and Complimentary Medicine [AMED], Excerpta Medica dataBASE [EMBASE], and PsycINFO) and Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases. The final search strategy included terms for web, health conditions, self-care, or behavior change. The final search strategies are available in the Multimedia Appendix 1. Corresponding authors were contacted for additional publications, including where only abstracts were located in the search. References of the included papers were screened for inclusion.

Study Selection

After deduplication, screening was performed in 2 stages. In stage 1, abstracts and titles were screened. In line with previous practice where a large number of studies were located, partial double screening with checks for accuracy were used [46]. A random 10% sample of the abstracts and titles were independently double screened for inclusion by ST and a single second researcher [47]. Agreement between reviewers of the 10% of titles and abstracts was good (87.5%, where the prevalence-adjusted and bias-adjusted kappa=0.75 indicates good agreement) [48]. Disagreement was resolved by discussion, and consensus and screening tools were refined in light of these. The remainder of the title and abstracts were reviewed by ST only. At stage 2, full texts were obtained and screened for inclusion by ST [49,50].

Data Extraction

One author (ST) extracted the data from the included studies. Where more than one outcome was reported in each category, the primary outcome was included. Quality of life (QoL) was categorized as a health or psychosocial outcome depending on whether there was a greater balance of health or disability or psychosocial questions in the QoL tool.

The PROGRESS Plus (PP) framework was used to identify the SDH that could contribute toward health inequalities in the included health conditions and in the context of web-based interventions [51]. The analyses that had explored the modification of intervention effectiveness by social characteristics in the PP framework were extracted.

Risk of Bias (Quality) Assessment

Methodological quality was assessed independently by ST and either PL, CC, or AH. Disagreement was resolved by discussion, and consensus was reached for all of the risk of bias (RoB) domains. The Cochrane collaboration RoB tool 1.0 was used to assess the quality of RCT studies [52]. The newly updated version 2.0 was not used as it does not allow for assessment of *other bias* and would therefore not allow us to capture issues with selective recruitment. The Risk Of Bias In Non-randomized Studies-of Interventions (ROBINS-I) tool was used to assess the quality of non-RCT studies [53]. These tools were used to produce an overall RoB rating for each study.

Much of the Cochrane RoB assessment is focused on ensuring that there is balance in the samples in the two arms of the study. However, the potential for selective recruitment is also important to ensure that the sample is representative of the population with the condition. Here, selection bias was assessed under the *other* category. Inclusion and exclusion criteria were examined to determine whether they potentially excluded people who experienced greater health inequity (eg, no access to the internet, not having the skills to use it, language barriers) and whether there was a discussion of the study population being representative of those with the condition.

Data Analysis and Synthesis

Descriptive tables were populated using the data from the included papers, accompanied by a narrative synthesis. A meta-analysis was not possible due to the heterogeneity of outcomes. A novel summary figure was developed for this systematic review, which was based on an adapted version of the Harvest plot, referred to here as the *Adapted Harvest plot* [54]. The Adapted Harvest plot allows for a direct comparison of the sample size of the studies where the effect was found (or not found) across outcomes, and an impression of the quality of the study through RoB. This gives an indication of the strength and validity of the findings in relation to each outcome. A key explanation of the features and representation of the Adapted Harvest plot is shown in Multimedia Appendix 2, and information is provided in each plot. To be inclusive, all reported trends (*P*<.10) were included as evidence regardless of whether they fell under the standard *P*<.05 probability cut off and were reported in the text.

Analysis of Subgroups or Subsets

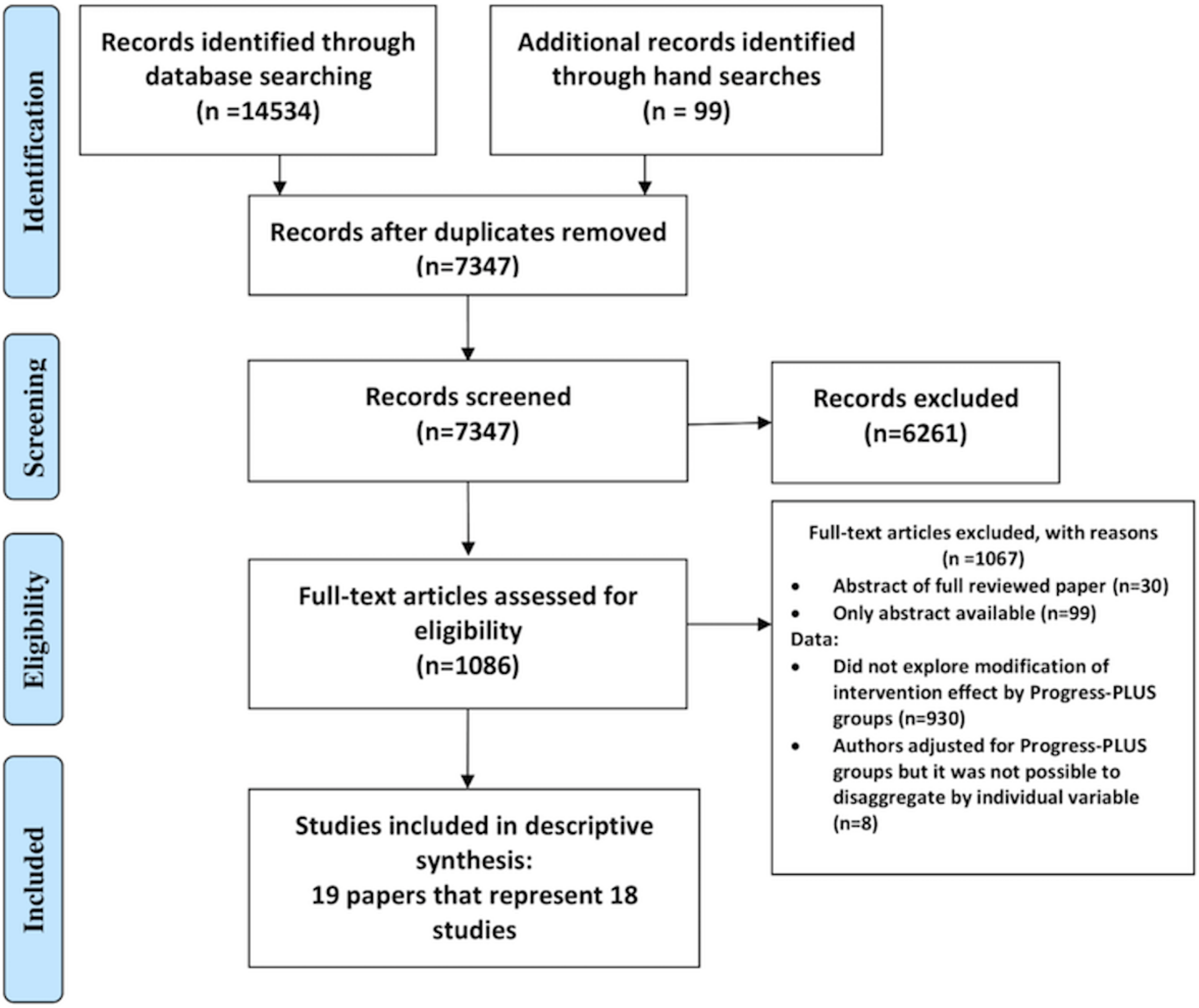
We intended to conduct subgroup analysis of differences in the application of behavioral change techniques and theory in intervention development and differences in use as potential mechanisms for modification of intervention effect by the PP group. However, these were inconsistently reported, making it challenging to analyze and draw any conclusions. Only 7 of the studies reported modification of use by PP characteristics, and 11 described the use of theory in intervention development. The selection and application of behavioral change techniques have largely not been described. Therefore, the exploration of mechanisms for modification of the intervention effect did not progress to a full analysis.

Results

Selection of Studies

After the removal of duplicates, 7346 records were obtained (Figure 1). A total of 18 studies (reported across 19 papers) were eligible for inclusion in the review (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis flow chart.



Description of Studies

Of the 18 included studies, 1 focused on asthma [55], 2 on COPD [56,57], 13 on diabetes (5 on type 1 [58-62], 6 on type 2 [63-68], and 2 on both [69,70]), and 2 studies included participants with osteoarthritis [71,72]. The study characteristics are described in Multimedia Appendix 2. In total, 9 of the studies were conducted in the United States, 3 in the Netherlands, 2 in Canada, and 1 in the United Kingdom, Australia, France, and Israel; 13 of the studies were RCTs and 5 were non-RCTs. In total, 6003 participants were included, and the study size ranged from 38 to 1799. The reported characteristics of the study participants are presented in Multimedia Appendix 3. Studies of both children and adults were included, with an age range of

12 to 75 years across the studies. Half of the participants were female (50.1%).

Only 1 study reported that the patients enrolled in the study were representative of the population with diabetes [69]. In total, 4 studies purposely recruited a highly diverse sample in terms of ethnicity [62,67,69,73]. Follow-up times varied from 2 weeks to 12 months, and the follow-up time point was not clear for 1 study [69]. Study attrition ranged from 0% [67,70] to 31% [65], and attrition was not clearly stated in 2 studies [69,71].

Intervention Content and Outcomes Targeted by the Intervention

In total, 4 of the studies explored the effectiveness of smartphone apps and the remaining 14 explored interventions delivered through websites. Descriptions of the intervention content are provided in Multimedia Appendix 4 and summarized below. In addition to the website or app, 3 interventions provided remote support by phone or video call [55,63,65,72], 1 provided a workbook [67], 1 provided a blood glucose monitor with wireless transfer [69], and 1 provided motivational interviewing before use of the website [61]. In 3 studies, there were 2 versions of the intervention, 1 with and 1 without external support (eg, email or phone support) [60,64,67].

Modification of Intervention Design for the Needs of Disadvantaged Social Groups

A total of 2 studies designed the intervention so it had an accessible format: one was designed to maximize usability for people with lower health literacy [65] and the other study used a serious-game intervention designed to be appealing to a range of ages (11-18 years) and a range of baseline knowledge levels in boys and girls [59]. Studies also designed interventions so they were appealing or accessible to ethnic minority groups. TEENCOPE was developed specifically for young people and used a graphic novel format and a cast of ethnically diverse characters with type 1 diabetes (T1D) who present challenging social situations, approaches to solving problems, and consequences of decisions [62]. Another study that targeted Latinos for a type 2 diabetes (T2D) self-management program was provided in Spanish and English [64,67,73].

Overall Duration of Intervention

Intervention duration varied from 4 weeks to 12 months, and 5 studies did not provide clear information on the duration [60,64,68,69,71].

Potential for Meta-Analysis

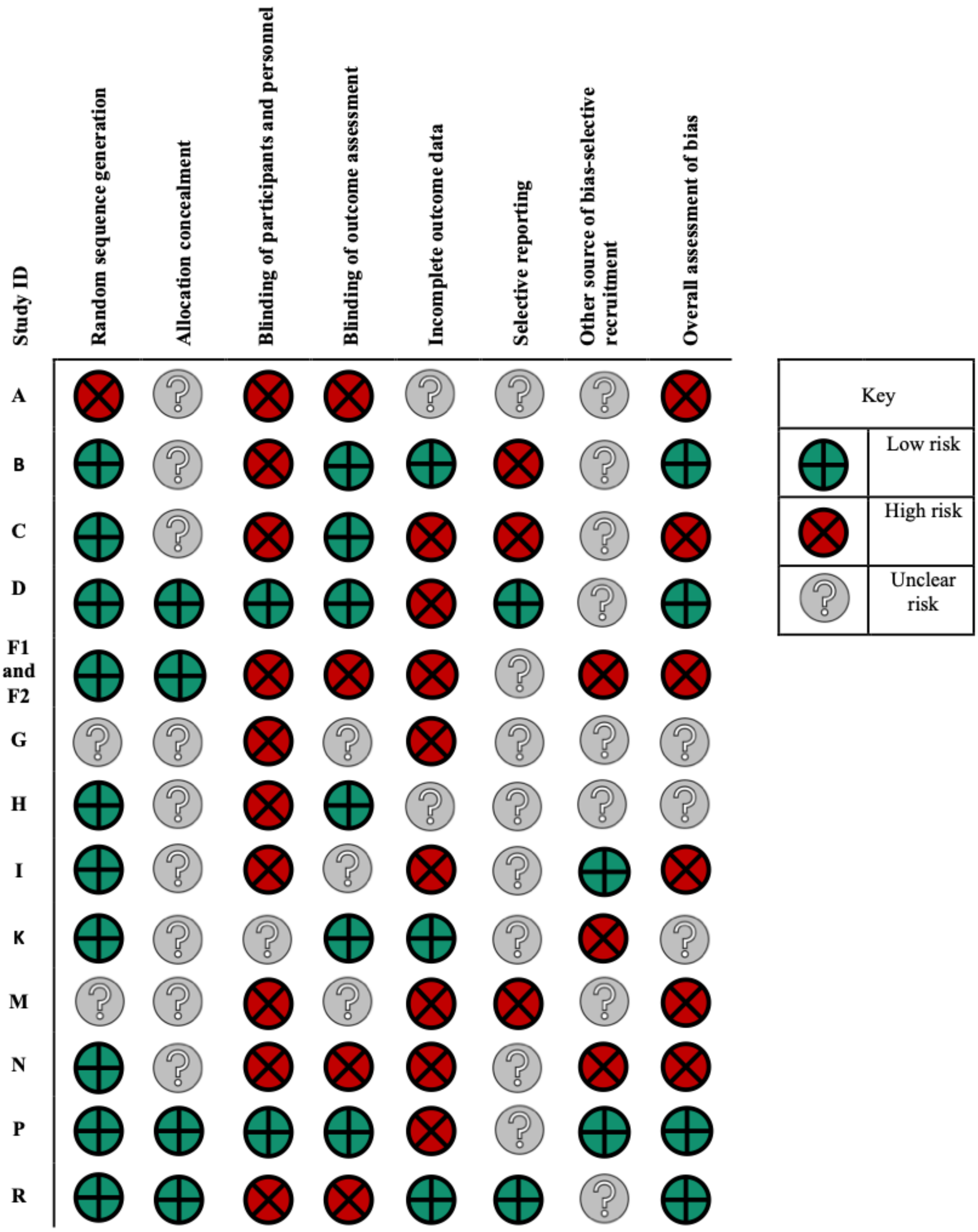
It was not possible to conduct meta-analysis due to differences in the outcomes and PP categories included in the modification of intervention effect analysis, resulting in a high level of heterogeneity. Narrative synthesis was used to present findings in relation to the research questions for each of the 4 health conditions.

Methodological Quality

Randomized Controlled Trial

Using the Cochrane RoB assessment, 4 of the 13 RCTs were considered to be low-risk RoB [56,62,63,72], 6 had high RoB [55,57,60,61,64,69], and 3 did not have enough information to assess RoB [58,66,67]. Methodological assessments for each of the domains in the Cochrane RoB assessment are provided in Figure 2. Overall, the lowest RoB came from the random sequence generation (low RoB in 10 studies, high in 1, and insufficient information was provided in 2 studies) and the highest came from blinding of participants and personnel (10 studies had high RoB, 2 had low, and 1 was unclear). However, given the nature of the digital interventions, it was often not possible to blind the participants and personnel. Selection bias was assessed under the *other* category, with 8 studies classified as unclear RoB, 3 as high, and 2 as low RoB.

Figure 2. Risk of bias table for randomized controlled trials.



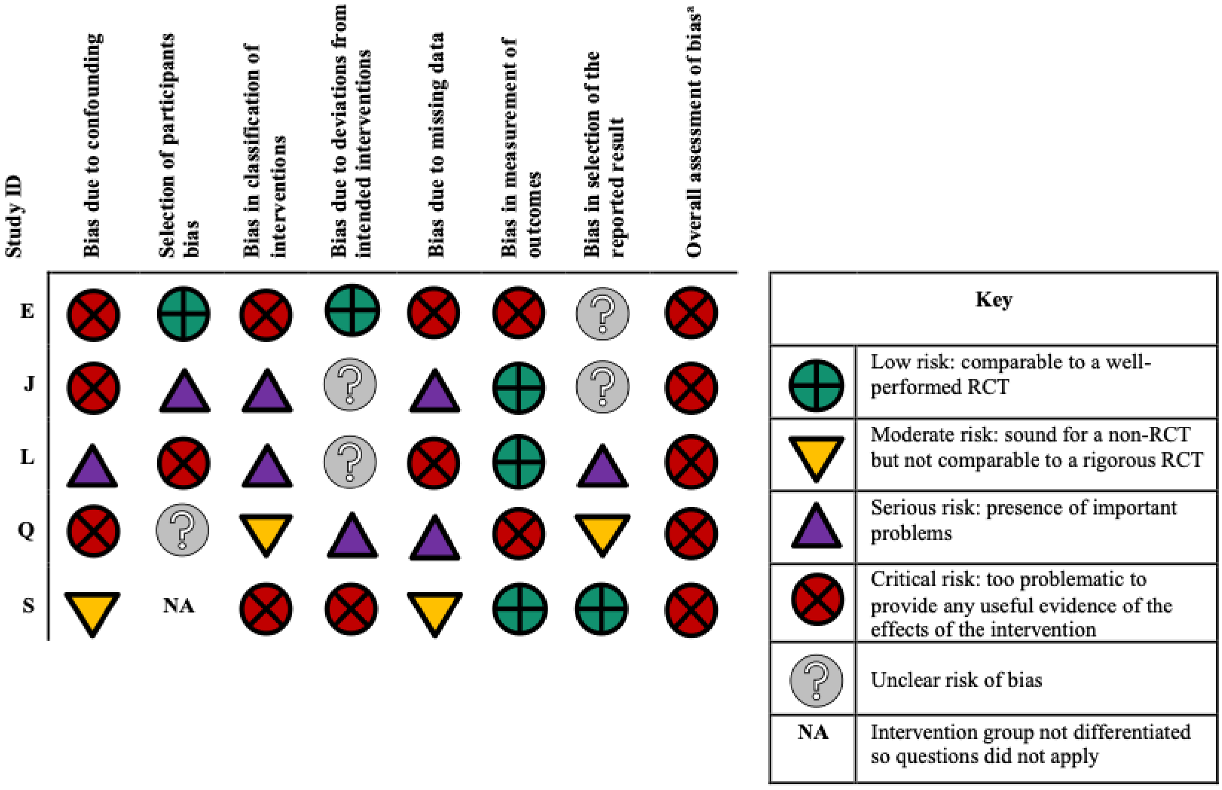
Study ID	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other source of bias-selective recruitment	Overall assessment of bias
A	⊗	?	⊗	⊗	?	?	?	⊗
B	⊕	?	⊗	⊕	⊕	⊗	?	⊕
C	⊕	?	⊗	⊕	⊗	⊗	?	⊗
D	⊕	⊕	⊕	⊕	⊗	⊕	?	⊕
F1 and F2	⊕	⊕	⊗	⊗	⊗	?	⊗	⊗
G	?	?	⊗	?	⊗	?	?	?
H	⊕	?	⊗	⊕	?	?	?	?
I	⊕	?	⊗	?	⊗	?	⊕	⊗
K	⊕	?	?	⊕	⊕	?	⊗	?
M	?	?	⊗	?	⊗	⊗	?	⊗
N	⊕	?	⊗	⊗	⊗	?	⊗	⊗
P	⊕	⊕	⊕	⊕	⊗	?	⊕	⊕
R	⊕	⊕	⊗	⊗	⊕	⊕	?	⊕

Key	
⊕	Low risk
⊗	High risk
?	Unclear risk

Nonrandomized Controlled Trial Studies

Using ROBINS-I, all 5 of the non-RCT studies were considered to have critical RoB (Figure 3) [59,65,68,70,71]. Across the studies, the lowest RoB came from the measurement of outcomes, and the highest RoB came from the classification of the intervention, and bias due to missing data.

Figure 3. Risk of bias assessment for the nonrandomized controlled trial (RCT) studies. [a]Overall risk of bias: equal to the most severe level of bias found in any domain.



Study ID	Bias due to confounding	Selection of participants bias	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall assessment of bias[a]
E	⊗	⊕	⊗	⊕	⊗	⊗	?	⊗
J	⊗	▲	▲	?	▲	⊕	?	⊗
L	▲	⊗	▲	?	⊗	⊕	▲	⊗
Q	⊗	?	▼	▲	▲	⊗	▼	⊗
S	▼	NA	⊗	⊗	▼	⊕	⊕	⊗

Key	
⊕	Low risk: comparable to a well-performed RCT
▼	Moderate risk: sound for a non-RCT but not comparable to a rigorous RCT
▲	Serious risk: presence of important problems
⊗	Critical risk: too problematic to provide any useful evidence of the effects of the intervention
?	Unclear risk of bias
NA	Intervention group not differentiated so questions did not apply

Modification of Intervention Effectiveness by PROGRESS Plus Groups

A total of 15 of the 18 studies that explored the modification of intervention effectiveness by PP categories reported effect modifiers. There was evidence that people were more likely to benefit from the intervention if they were from a minority ethnic group, were employed, had a higher level of education, and had divorced parents (study of adolescents). The findings for age, gender, and health literacy were mixed. There was no evidence of an interaction income, numeracy, or the number of people living in the household. The full key to the Adapted Harvest plot is shown in Table 1. Further details including outcomes and estimates (where provided) where interactions were found are presented in Multimedia Appendix 5 and a matrix table containing an overview of interactions across the conditions and PP groups is provided in in the Multimedia Appendix 6.

Table 1. Key to the Adapted Harvest plot.

Feature	Representation
Direction of effect category	- Positive: favors the PROGRESS Plus group ($P<.10$) - No effect: study found no evidence of an effect ($P\geq.10$) - Negative: favors the comparator group ($P>.10$)
Stack height	- Study size
Stack color	- Risk of bias assessment - RCT[a] studies: Low risk—blue, high risk—purple, unclear—gray - Non-RCT studies: low risk—blue, moderate risk—yellow, serious—orange, critical—purple, not enough information—gray
Stack pattern	- RCT studies: solid colors - Non-RCTs: patterned with dots
Number within the stack	- Study ID
Bar size	- Total number of participants in the studies finding evidence of a positive association, no effect, or a negative association with the outcome

[a] RCT: randomized controlled trial.

Age

A total of 14 studies (1 asthma, 2 COPD, 9 diabetes, and 2 osteoarthritis) examined the modifying effect of age on at least one outcome, with evidence of mixed effects (younger and older benefited more) across asthma, COPD, and diabetes studies.

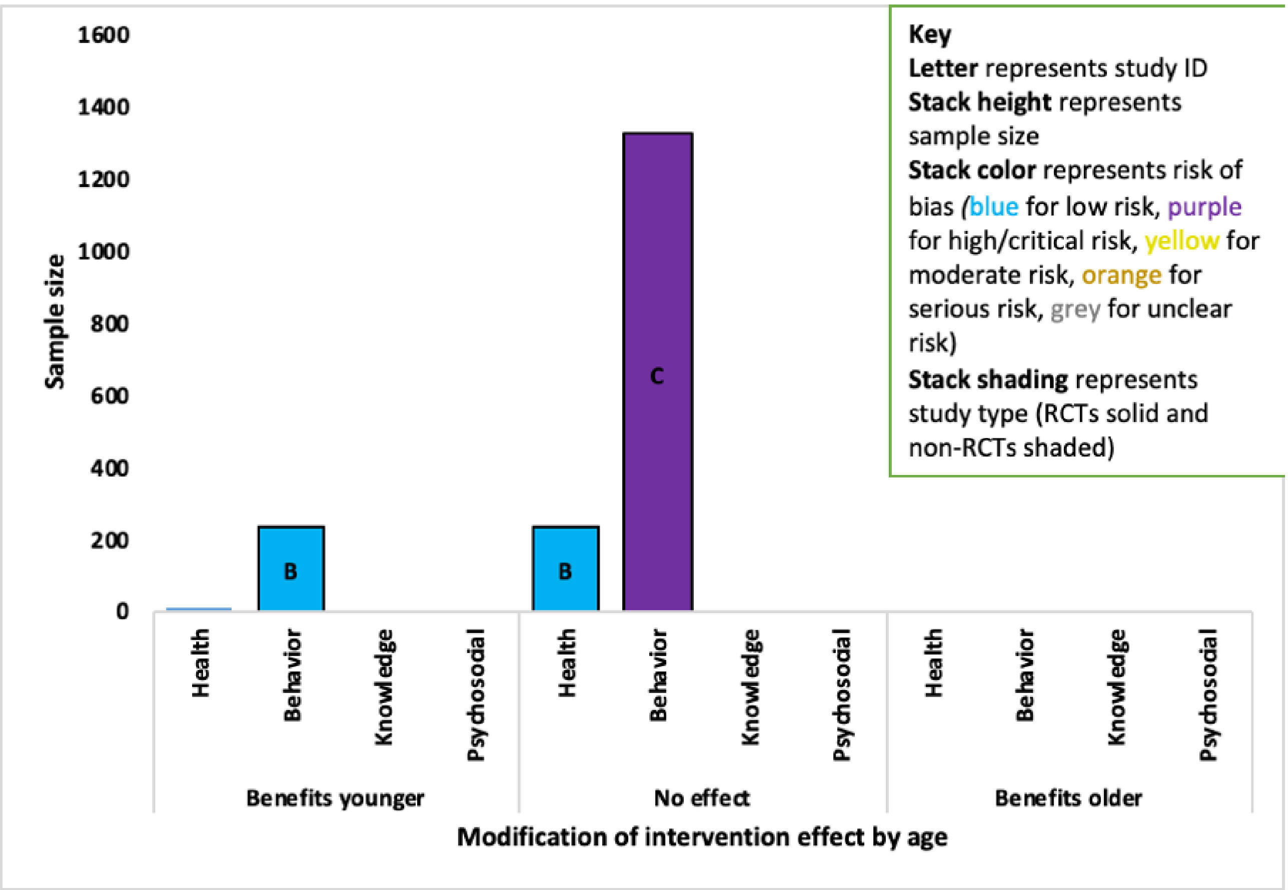
Asthma

The one asthma study (ID A; n=234; high RoB) found that increasing age was associated with increased medication adherence among adolescents aged 12 to 18 years [55]. Those scoring in the higher range in the Medication Adherence Report Scale (>19) were on average aged 0.7 years ($P=.02$) than those who scored in the lower range.

Chronic Obstructive Pulmonary Disease

One of the 2 COPD studies found evidence that older participants benefited less from the intervention than younger people on the behavioral outcome (Figure 4). The higher quality evidence came from an RCT (ID B; n=239; low RoB) that indicated that a 1-year increase in age was associated with a 33-point decrease in change in daily step count ($P=.03$) but found no association with the health-related QoL outcome [56]. The evidence for no effect on the behavioral outcome came from an RCT (ID C) with a high RoB but a larger sample (n=1325) [57].

Figure 4. Adapted Harvest plot with evidence for the modification of intervention effect by increasing age for chronic obstructive pulmonary disorder studies. RCT: randomized controlled trial.

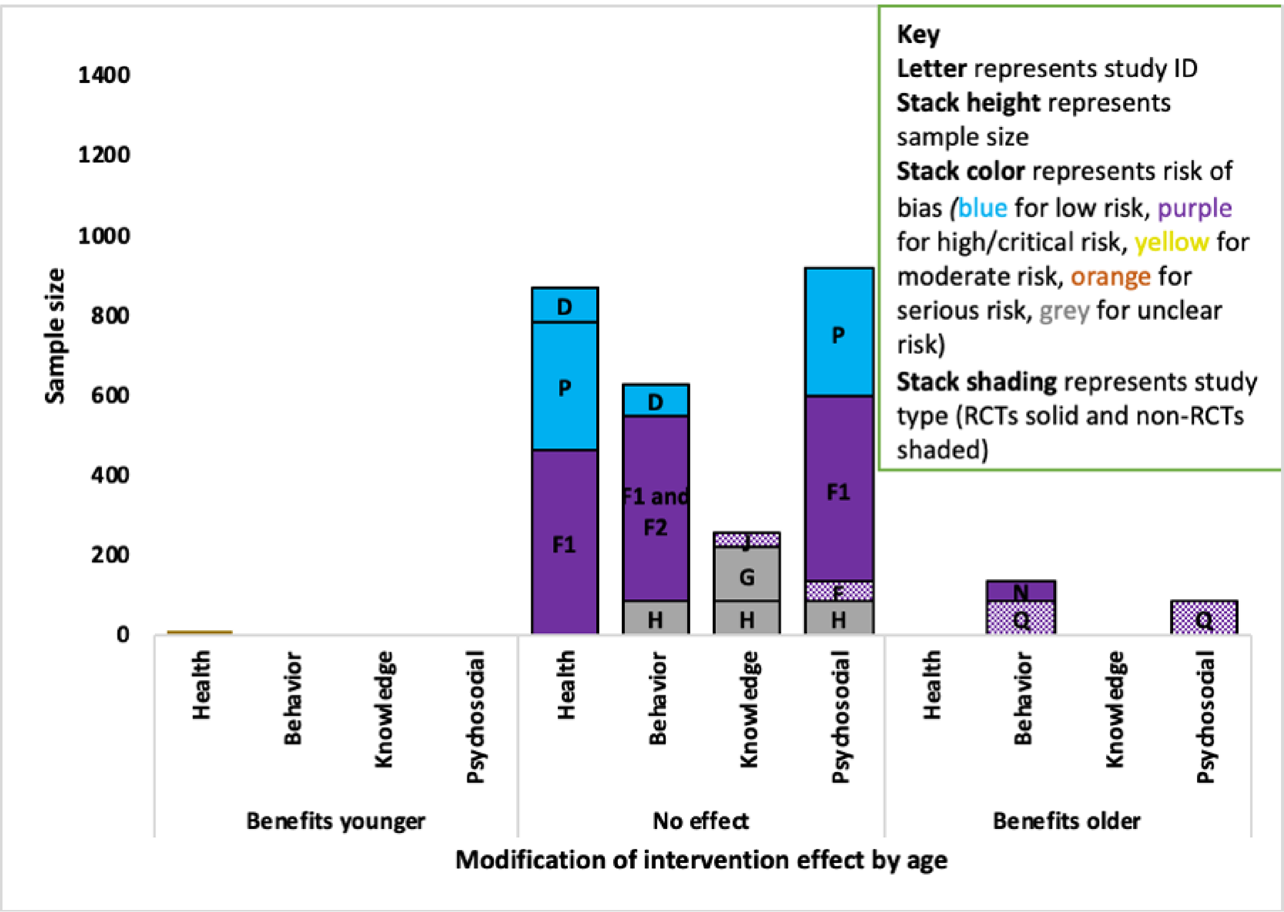


Diabetes

Of the 9 diabetes studies that reported the modification of intervention effectiveness by age, 2 studies indicated that older participants benefited more from using the intervention (Figure 5).

For the behavioral outcome, the strongest evidence indicated that there was no modification of intervention effect by age, which came from 3 studies (IDs D, F, and H) with an overall sample of 628 patients and a low, high, and unclear RoB [58,63,64,73]. There was evidence that older participants benefited more in 2 small studies (combined n=133) with a high and critical RoB: a non-RCT (ID Q; n=81; critical RoB) found that older participants benefited more (*P*<.001) in adults aged ≥25 years with T2D [68], and an RCT (ID N; n=52; high RoB) found older adolescents (aged 16-18 years) improved more (*P*<.01) than younger adolescents with T1D (aged 13-18 years) [61].Regarding psychosocial outcomes, the strongest evidence indicated that there was no modification of effect by age and it came from 4 studies (IDs P, F, E, and H, combined sample n=915) with low, unclear, high, and critical RoB [58,62,64,65]. Whereas evidence for modification for intervention effect on the behavioral outcome (*P*=.01) in adults aged ≥25 years with T2D came from a small non-RCT (ID Q; n=81; critical RoB) [68]. There was no evidence of an interaction effect with age across the health outcomes in 3 studies (IDs D, P, and F) [62-64] or with diabetes knowledge in 3 studies (IDs J, G, and H) [58,59,66].

Figure 5. Adapted Harvest plot with evidence for increasing age-modifying intervention effects across outcomes in diabetes studies. RCT: randomized controlled trial.



Osteoarthritis

There was no evidence of a difference in effectiveness by age for health outcomes in the 2 studies [71,72].

Gender

A total of 12 studies (8 diabetes, 1 asthma, 1 COPD, 2 osteoarthritis) examined the modifying effect of sex on at least one outcome, with mixed findings in diabetes studies.

Asthma, Chronic Obstructive Pulmonary Disease, and Osteoarthritis

There was no evidence of a difference in effectiveness by gender on a behavioral outcome (medication adherence) in the asthma study (ID A) [55], a behavioral outcome in the COPD study (ID C) [57], or on health outcomes in either of the osteoarthritis studies (IDs R and S) [71,72].

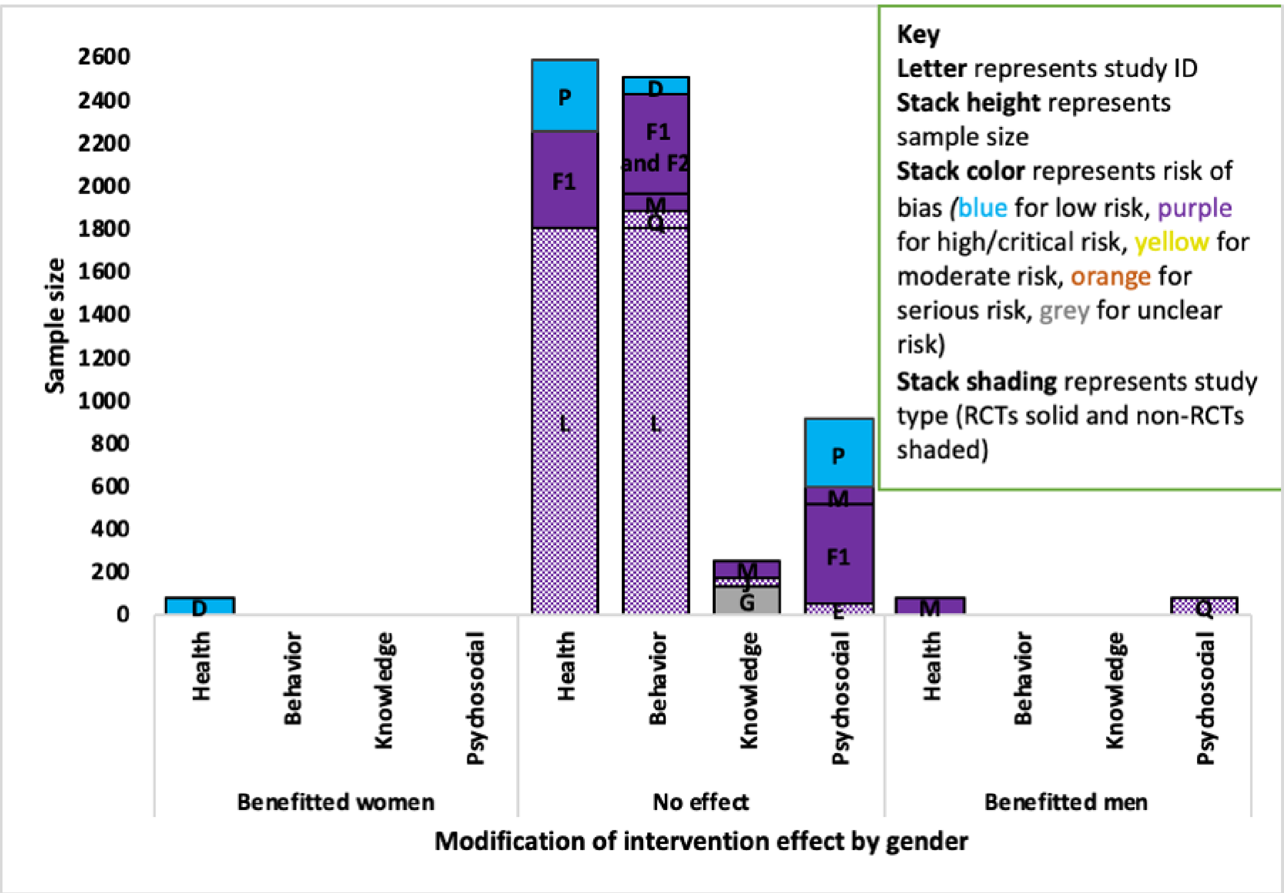
Diabetes

Of the 9 diabetes studies that explored the modification of intervention effect by gender, 3 small studies (IDs D, M, and Q) found evidence of a difference. Two studies (IDs M and Q) indicated that male participants benefited more, and one study (ID D) indicated that female participants benefited more (Figure 6).

The evidence that gender modified the health outcome was mixed. The strongest evidence indicated that gender did not modify the intervention effect and it came from 3 studies (IDs P, F1, and L; combined n=2582) with low, high, and critical RoB [62,64,70]. Of the studies that indicated gender-modified intervention effectiveness, one small RCT (ID D; n=84) with a low RoB indicated female participants benefited more (P=.03) [63] and a small RCT (ID M; n=79) with a high RoB indicated male participants benefited more (P=.06) [60].

For psychosocial outcomes, all of the evidence came from studies of high and critical RoB. The evidence that male participants benefited more (P=.01) on psychosocial outcomes came from one small non-RCT (ID Q; n=81) with a high RoB [68], in comparison with 4 studies (IDs P, M, F1, and E) that found no evidence of an effect with a combined sample of 862 and low, high, critical, or unclear RoB (Figure 6) [60,64,65]. There was no evidence of the intervention effect being modified by gender for the behavioral outcomes in 5 studies (IDs D, F, M, Q, and L) or knowledge outcomes in 3 studies (IDs M, J, and G) [60,64,66,68,73].

Figure 6. Adapted Harvest plot for gender-modifying intervention effect across outcomes in diabetes studies. RCT: randomized controlled trial.



Education

A total of 7 studies (5 diabetes, 1 COPD, and 1 osteoarthritis) examined the modifying effect of education as an outcome, with evidence that higher education benefited more in 1 diabetes study.

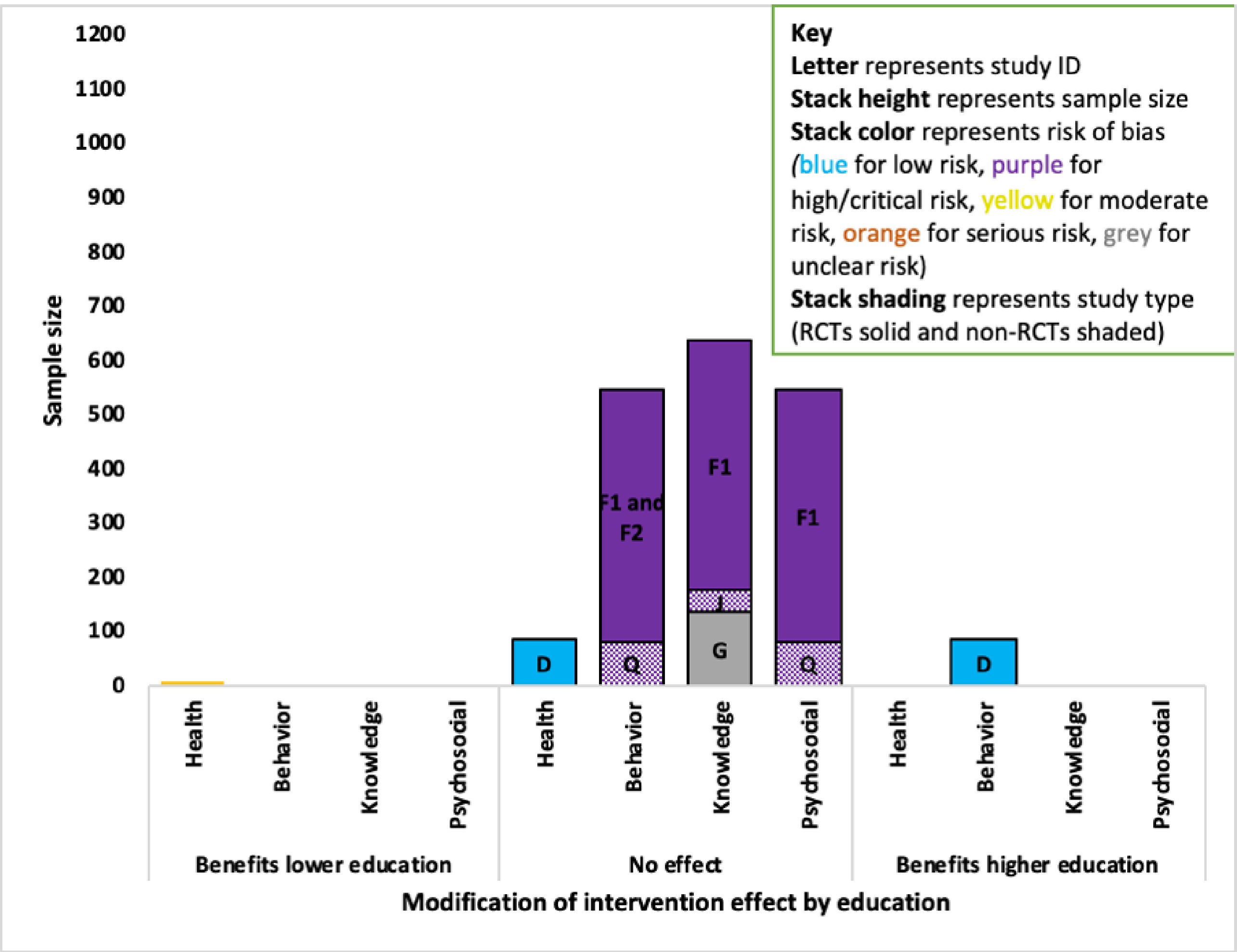
Chronic Obstructive Pulmonary Disease and Osteoarthritis

There was no evidence that the levels of education-modified intervention effectiveness in one COPD (ID C) study on a behavioral outcome [57], or an osteoarthritis study (ID R) on health outcomes [72].

Diabetes

There was evidence that those with higher education benefited more in one study, and no evidence of a difference in 4 diabetes studies (Figure 7). The strongest evidence came from a small RCT (ID D; n=84) with low RoB, which indicated that those with higher education benefited more from the intervention on a behavioral outcome (*P*=.03) but did not find an effect on a heath outcome [63]. The combined sample of the 2 studies where there was no effect on the behavioral outcome was 544, with a high (ID F1 and F2; n=463) [64,73] and critical RoB (ID Q; n=81) [68]. There was no evidence that education-modified intervention effect on knowledge outcomes in 3 studies (IDs F1, J, and G) [59,64,66] or psychosocial outcomes in 2 studies (IDs F1 and Q) [64,68].

Figure 7. Adapted Harvest plot for higher education modifying intervention effects across outcomes in diabetes studies. RCT: randomized controlled trial.



Ethnicity

A total of 8 studies (7 diabetes and 1 osteoarthritis) examined the modifying effect of ethnicity on study outcomes, with evidence that ethnic minority groups benefited more from the interventions in diabetes studies.

Diabetes

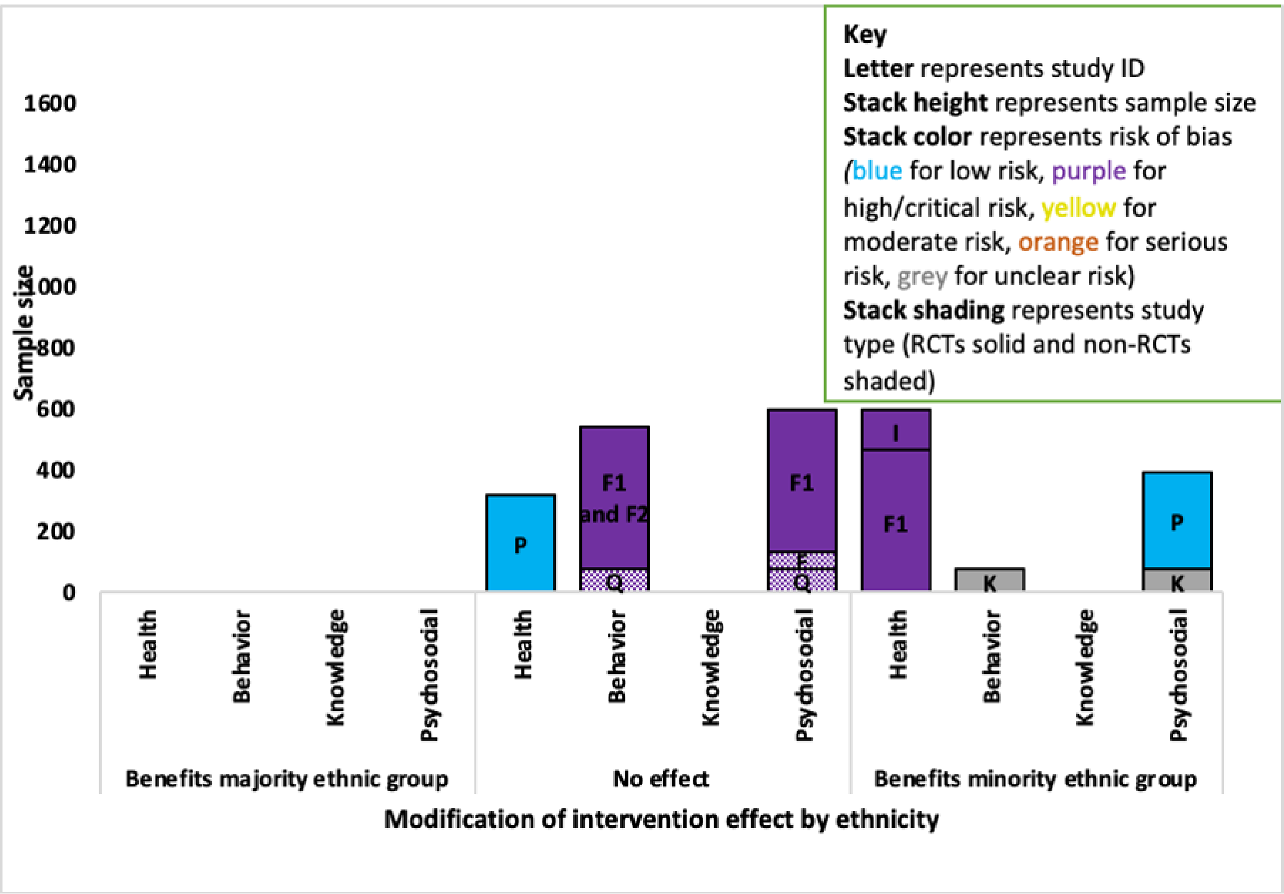
There was evidence that minority ethnic groups benefited more from the intervention than majority ethnic groups in 4 of the 7 diabetes studies that explored this interaction (Figure 8).

For the health outcomes, an RCT study with a low RoB (ID P; n=320) found no evidence that the intervention effect was modified by ethnicity [62], while 2 RCTs (ID F1 and I; combined sample n=597) both with a high RoB indicated that people from minority ethnic groups benefited more on the health outcome (ID F1; *P*=.006; for study ID I no estimates were provided) [64,69].

Regarding the behavioral outcome, a small RCT (ID K; n=73) with an unclear RoB found that minority ethnic groups benefited more (*P*=.01) [67], and the evidence of no effect came from one non-RCT (ID Q; critical RoB) and an RCT (ID F; high RoB) with a combined sample of 544 [64,68,73].

For the psychosocial outcomes, evidence indicating that minority groups benefited more from the intervention came from 2 RCTs (combined sample n=393): ID P with a low RoB (n=320; *P*=.07) [62] and ID K (n=73; *P*=.003) with an unclear RoB [67]. The evidence of no effect on the psychosocial outcome came from 3 studies (IDs F1, E, and Q; combined sample n=595): an RCT with a high RoB and 2 non-RCTs with critical RoB [64,65,68]. No studies reported the modification of a knowledge outcome by ethnicity.

Figure 8. Adapted Harvest plot for minority ethnic group modifying intervention effects across outcomes in diabetes studies. RCT: randomized controlled trial.



Osteoarthritis

There was no evidence that ethnicity-modified intervention effect on a health outcome in adults with osteoarthritis in a non-RCT (ID S) [71].

Employment

A total of 3 studies (1 COPD, 1 diabetes, and 1 osteoarthritis) examined the modifying effect of employment on at least one outcome, with evidence that employed participants benefited more in an osteoarthritis study.

Chronic Obstructive Pulmonary Disease and Diabetes

There was no evidence that employment was a moderator of intervention effectiveness in one RCT COPD study (ID C) [57] or one non-RCT diabetes study (ID Q) [68].

Osteoarthritis

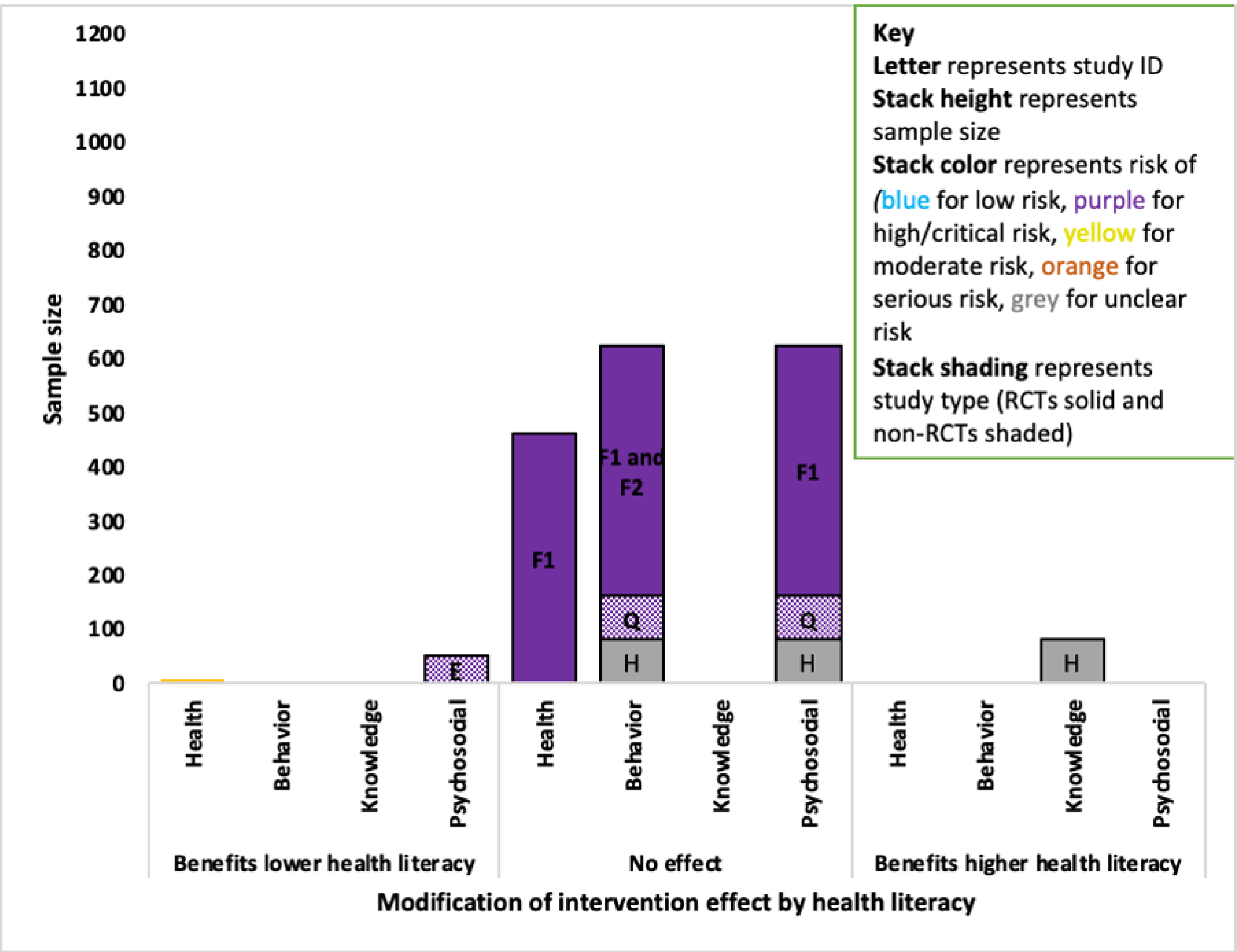
An osteoarthritis RCT with a low RoB (ID R; n=148) found that participants who were employed showed greater improvements in health outcomes (walking pain) than unemployed participants 3 months after using the intervention (interaction: *P*=.02) [72].

Health Literacy

A total of 4 diabetes studies examined the modifying effect of health literacy on study outcomes, with 2 of the 4 diabetes studies reporting evidence of a difference. However, the 2 studies provided evidence in different directions for different outcomes (Figure 9). The evidence from both studies was weak: a small RCT (ID H; n=81) with an unclear RoB found evidence that people with higher health literacy benefited more from the intervention on a knowledge outcome (no estimates provided) [58].

For psychosocial outcomes, 3 studies (combined sample: n=625) indicated that health literacy did not modify the intervention effect; these were 2 RCTs with an unclear (ID H) and high RoB (ID F1) and a non-RCT with a critical RoB (ID Q) [58,64,68]. A small single-arm pilot non-RCT (ID E; n=51) with a critical RoB found that those with lower health literacy benefited more on psychosocial outcomes (*P*=.02) [65]. There was no evidence that health literacy modified the intervention effect on a health outcome in 1 study (ID F1) [64] or behavior change outcomes in 3 studies (IDs F, Q, and H) [58,64,68,73].

Figure 9. Adapted Harvest plot for higher health literacy modifying intervention effects across outcomes in diabetes studies. RCT: randomized controlled trial.



Family Structure

One small non-RCT (ID J; n=38) diabetes study with a critical RoB examined the modifying effect of family structure on study outcomes, finding children of divorced parents benefited more from the intervention. A higher proportion of children (aged 11-18 years) with T1D who were in the high-effect subgroup for change in diabetes knowledge (benefited more from the intervention) had divorced parents (35%) compared with the low-effect subgroup (6%; *P*=.03) [59].

Income, Numeracy, Number of People Living in the Household

There was no evidence that income modified effectiveness on 2 diabetes studies (1 non-RCT: ID Q and 1 RCT: ID P) [62,68] or numeracy in 1 RCT (ID F) [64]. There was no evidence that the number of people living in the household modified the effectiveness of an intervention on health outcomes in a non-RCT osteoarthritis study (ID S) [71].

Discussion

Principal Findings

This review examined the modifying effects of participant characteristics on the effectiveness of web-based interventions for chronic health conditions. We found evidence that intervention effectiveness was modified by participants' social characteristics. In the diabetes literature, there was evidence that people from minority ethnic groups gained greater benefits from interventions than majority ethnic groups. There was evidence from single studies with variable quality that those with higher education, divorced parents (adolescent), and who were employed benefited more from interventions. A small high-quality diabetes study indicated that those with a higher level of education benefited more from the intervention. A small low-quality diabetes study found that a higher proportion of adolescents with divorced parents (adolescents) were in the high-effect intervention group compared with the low-effect group. A high-quality osteoarthritis study indicated that employed participants were more likely to benefit from the intervention than unemployed participants. The findings for modification of the effect by participant age were mixed. Older people were found to benefit less from the intervention in high-quality COPD studies, and they were found to benefit more from the intervention in 2 low-quality diabetes studies and a low-quality asthma study. Gender and health literacy were only considered in studies on diabetes and showed mixed effects. A total of 2 small low-quality studies indicated that male participants benefited more from diabetes interventions, and 1 small high-quality study indicated that female participants benefited more. Two studies had contradicting findings for health literacy: a small study with an unclear RoB indicated that those with higher health literacy benefited more, while a small low-quality study found that those with lower health literacy benefited more. There was no evidence of a modification to intervention effectiveness by income, numeracy, or the number of people living in the household.

The strength of evidence across the studies could not be assessed through a meta-analysis as not all studies provided estimates, and the outcomes were heterogeneous. Therefore, the strength of the evidence was explored through study size, RoB, and estimates, where possible. The majority of the evidence was drawn from studies with high and uncertain RoBs.

Strengths and Limitations of the Methods Used

To our knowledge, this is the only systematic review that has investigated whether there are equal benefits of web-based self-care interventions for people with different characteristics. The breadth of the studies reviewed here is both a strength and a weakness. The inclusion of 4 different physical health conditions meant that it was not possible to combine the evidence in cases where the outcomes targeted were heterogeneous, which limits what can be inferred from the results. However, we summarized the available evidence, providing the first systematic exploration of how PP characteristics modify intervention effectiveness. Simple vote counting was not judged appropriate for the synthesis of findings that could not be meta-analyzed. Instead, we created a novel summary figure based on the Harvest plot referred to here as the *Adapted Harvest plot*. The Adapted Harvest plot provided an indication of the strength of the evidence for narrative synthesis by including study size and RoB.

The majority of the screening was conducted by one person, and only 10% of the abstracts and titles were double screened. This is in line with previous practice where a large number of studies were located [46,74-77]. Every effort was made to locate all relevant literature; however, it is possible that some relevant studies may have been missed.

Limitations of the Evidence Base

There were several limitations in the methodology and data reported by the studies included in the review. It was not possible to conduct a meta-analysis because the outcomes were heterogeneous, and essential data were not reported. A high proportion of the included studies found that the intervention effect was modified by at least one participant characteristic on at least one outcome. This suggests that teams may be more likely to publish these analyses when they find evidence of a difference in effect [78].

There was a high risk of selection bias across the included studies, which may have excluded people from lower SES groups. The few studies that did comment on the representativeness of the study population indicated that the samples tended to be more white people, with higher levels of education. This not only limits the generalizability of the study findings but also potentially masks differences in effectiveness that may have been present between more and less advantaged groups because the sample is underpowered to detect differences in these subgroups.

The complex relationship between social characteristics and potential effectiveness or engagement with online interventions was not considered carefully in the included studies. The range of different PP characteristics explored and the comparisons within the PP characteristic appeared to be decided post hoc and did not appear to be guided by theory. Therefore, the evidence on the influence of PP on effectiveness was limited.

The Cochrane RoB 1.0 tool was limited in evaluating bias in the data that were important for this systematic review. Although the tool is very effective at identifying bias that can arise from the arms of a study with an unbalanced sample, it does not provide a category that evaluates the risk of the sample not being representative of the general population with the condition. RoB 2.0 was not used in this study because it does not allow for *other* RoBs and would therefore not allow us to capture issues with selective recruitment.

Comparison With Prior Work

There is no previous systematic evidence comparing the health equity effects of web-based self-care interventions for people with different social characteristics. This review agrees with previous evidence from single studies that have found that web-based self-care interventions can benefit underserved and disadvantaged groups when the intervention has been designed specifically for such groups [14,25,26]. Some of the included studies that found evidence that underserved groups benefited more from the intervention had modified their interventions to be more accessible, useable, or engaging for these groups. Two studies designed the intervention, so it was in an accessible format for those with different educational abilities. A study found that those with lower health literacy benefited more and designed interventions to maximize usability for people with lower literacy [65]. A study found that those with divorced parents and lower baseline knowledge benefited more from the intervention and opted for a serious-game intervention that was designed to be appealing to a range of ages (11-18 years) and therefore a range of baseline knowledge in boys and girls [59].

Of the 4 studies that found that ethnic minority groups benefited more, 2 studies had adapted the intervention to be appealing to the study population targeted. One study targeting Latinos with T2D provided the intervention in Spanish and English [64,73]. Another study that found that ethnic minority youth with T1D benefited more from using the intervention than majority ethnic groups used an intervention with a graphic novel format and a cast of ethnically diverse characters with T1D [62]. However, there are no specific design features that suggest why ethnic minority groups may benefit more than majority ethnic groups in these studies. Indeed, the other 2 studies that found ethnic minority groups benefited more did not adapt the interventions to make it more accessible to ethnic minority groups. It is possible that in these studies, ethnic minority groups may have benefited more from the intervention because they had less exposure to health care support before using the digital intervention. There has been previous evidence that those from ethnic minority groups and those with lower SES face greater challenges accessing health care services and support [79-81]. Therefore, interventions that reduce barriers to access and use may be more effective for populations currently underserved by health care services.

Common to the 4 studies that found minority ethnic groups benefited more from the intervention was their sampling strategy aimed at maximizing the recruitment of minority ethnic groups [62,67,69,73]. Subsequently, all 4 studies had a high representation of people from ethnic minority groups in the sample, resulting in the sample being powered to detect differences in effectiveness by ethnicity. Webb et al [82] similarly found that recruitment sampling was an important predictor of the effectiveness of an intervention. They found that when theory or predictors were used to select recipients for the intervention, the intervention had the greatest improvements in behaviors [82]. The 4 studies in this review cited the potential for digital interventions to increase access to health care in these minority groups, as motivation for their study design, and target the recruitment of minority groups [62,67,69,73]. Therefore, consideration of the sample where the intervention was evaluated appears to be important in addition to considering the needs of the target population.

Interventions designed without considering the needs of the users can exclude social groups, and this is the likely cause of the difference in effectiveness found by education and employment. Van Dijk's theory of the digital divide proposed that if the content of the technology only fulfils the needs of the dominant group (eg, high education, employed) or is challenging to use, those users from the less dominant group will benefit less from the use of the technology [83]. This supposition has been supported by findings that the design of web-based health information can limit the usability of digital interventions for people of lower SES [36,37,84,85].

The mixed findings for age, gender, and health literacy may be associated with whether the interventions were designed considering the needs of people with those characteristics. This was illustrated by the findings for health literacy in this review. Davis et al [65] designed their intervention with the needs of low literacy individuals in mind and found that those with lower literacy levels benefited more from the intervention. Huang et al [58] found that those with higher literacy benefited more and acknowledged that they would need to provide additional support for users with lower health literacy levels at baseline. Alternatively, mixed findings may be related to the participant's other social characteristics. For example, in studies where older people were found to benefit more from the intervention, they may have had a higher level of education and consequently higher digital skills relative to their younger counterparts. There is growing evidence that individual social characteristics do not work in isolation but interact in complex ways that influence health outcomes [86]. As such, conducting an analysis involving the comparison of individual groups may not be sufficient to establish how digital self-care interventions may impact health inequities.

Conclusions

There was evidence that web-based self-care interventions for chronic conditions can benefit some (minority ethnic groups, divorced parents) and disadvantage other (low education, unemployed) social groups who have historically experienced health inequity. However, these findings should be treated with caution as most of the evidence came from a small number of low-quality studies. The findings for gender and health literacy were mixed across diabetes studies, and the findings for age were mixed across asthma, COPD, and diabetes studies. There was no evidence that income, numeracy, or the number of people living in the household modified intervention effectiveness. We conclude that there appear to be interaction effects that warrant exploration in future research, and a priori consideration of predicted interaction effects is recommended.

Acknowledgments

This project was funded by the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR; project reference 2019-5007) at the University of Bristol. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

Authors' Contributions

ST had primary responsibility for conducting this study as part of her NIHR SPCR PhD and drafted the manuscript. All authors contributed to the conception, study design, study quality assessment, analysis, and interpretation. All authors contributed to revisions of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Systematic review search strategies.
[DOCX File , 37 KB - jmir_v22i6e17849_app1.docx]

Multimedia Appendix 2
Table 2. Characteristics of included studies and populations.
[DOCX File , 28 KB - jmir_v22i6e17849_app2.docx]

Multimedia Appendix 3
Table 3. Participant characteristics.
[DOCX File , 26 KB - jmir_v22i6e17849_app3.docx]

Multimedia Appendix 4
Table 4. Intervention description.
[DOCX File , 34 KB - jmir_v22i6e17849_app4.docx]

Multimedia Appendix 5
Table 5: Overview of available data for social characteristics modifying intervention effects.
[DOCX File , 23 KB - jmir_v22i6e17849_app5.docx]

Multimedia Appendix 6
Table 6. Modification of effectiveness data available for PROGRESS-Plus categories in the included studies.
[DOCX File , 23 KB - jmir_v22i6e17849_app6.docx]

References

1. World Health Organization. Health Systems: Equity URL: http://www.who.int/healthsystems/topics/equity/en/ [accessed 2018-08-03]
2. Department of Health. Government of UK. 2012. Long Term Conditions Compendium of Information: Third Edition URL: https://www.gov.uk/government/publications/long-term-conditions-compendium-of-information-third-edition [accessed 2020-04-23]
3. World Health Organization. Chronic Diseases and Health Promotion. Chronic Diseases and Health Promotion: Integrated Chronic Disease Prevention and Control URL: http://www.who.int/chp/about/integrated_cd/en/index2.html [accessed 2018-05-30]
4. Epping-Jordan JE, Pruitt SD, Bengoa R, Wagner EH. Improving the quality of health care for chronic conditions. Qual Saf Health Care 2004 Aug;13(4):299-305 [FREE Full text] [doi: 10.1136/qhc.13.4.299] [Medline: 15289634]
5. Østbye T, Yarnall KS, Krause KM, Pollak KI, Gradison M, Michener JL. Is there time for management of patients with chronic diseases in primary care? Ann Fam Med 2005;3(3):209-214 [FREE Full text] [doi: 10.1370/afm.310] [Medline: 15928223]
6. Iacobucci G. CCG criticises NHS England after being ordered to cut GP funding. Br Med J 2018 Jan 17;360:k247. [doi: 10.1136/bmj.k247] [Medline: 29343493]
7. British Medical Association. 2016. Urgent Prescription for General Practice URL: https://archive.bma.org.uk/collective-voice/influence/key-negotiations/training-and-workforce/urgent-prescription-for-general-practice [accessed 2020-04-23]
8. Furler J, Harris M, Rogers A. Equity and long-term condition self-management. Chronic Illn 2011 Mar;7(1):3-5. [doi: 10.1177/1742395310386978] [Medline: 21398320]
9. Data From the Government of UK. 2006. General Household Survey URL: https://data.gov.uk/dataset/138ca035-a90c-4e37-80f5-4c73eeb6ae04/general-household-survey; [accessed 2020-03-10]
10. Gama-e-Colombo D. Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the commission on social determinants of health. Rev Direito Sanit 2010 Feb 1;10(3):253-266. [doi: 10.11606/issn.2316-9044.v10i3p253-266]
11. World Health Organization. 2016. Social Determinants of Health URL: http://www.who.int/social_determinants/en/ [accessed 2016-08-18]
12. Imison C, Naylor C, Buck D, Curry N, Smithson R, Ross S, et al. The King's Fund: Ideas That Change Health and Care. 2015. Transforming Our Health Care System: Ten Priorities for Commissioners URL: https://www.kingsfund.org.uk/publications/articles/transforming-our-health-care-system-ten-priorities-commissioners [accessed 2020-04-23]
13. Pekmezi DW, Williams DM, Dunsiger S, Jennings EG, Lewis BA, Jakicic JM, et al. Feasibility of using computer-tailored and internet-based interventions to promote physical activity in underserved populations. Telemed J E Health 2010 May;16(4):498-503 [FREE Full text] [doi: 10.1089/tmj.2009.0135] [Medline: 20507203]
14. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. Int J Med Inform 2002 Nov 12;65(3):169-177. [doi: 10.1016/s1386-5056(02)00048-5] [Medline: 12414016]
15. Safran C. The collaborative edge: patient empowerment for vulnerable populations. Int J Med Inform 2003 Mar;69(2-3):185-190. [doi: 10.1016/s1386-5056(02)00130-2] [Medline: 12810122]
16. Starren J, Hripcsak G, Sengupta S, Abbruscato CR, Knudson PE, Weinstock RS, et al. Columbia university's informatics for diabetes education and telemedicine (IDEATel) project: technical implementation. J Am Med Inform Assoc 2002;9(1):25-36 [FREE Full text] [doi: 10.1136/jamia.2002.0090025] [Medline: 11751801]

17. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
18. National Health Service. NHS Long Term Plan URL: https://www.england.nhs.uk/long-term-plan/ [accessed 2020-04-23]
19. Wilton Park. 2018. Reimagining Global Health: Self-Care Interventions and Implications for Healthcare (WP1639) URL: https://www.wiltonpark.org.uk/event/wp1639/ [accessed 2020-03-10]
20. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999 Jan;37(1):5-14. [doi: 10.1097/00005650-199901000-00003] [Medline: 10413387]
21. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
22. Franek J. Self-management support interventions for persons with chronic disease: an evidence-based analysis. Ont Health Technol Assess Ser 2013;13(9):1-60 [FREE Full text] [Medline: 24194800]
23. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
24. Chodosh J, Morton SC, Mojica W, Maglione M, Suttorp MJ, Hilton L, et al. Meta-analysis: chronic disease self-management programs for older adults. Ann Intern Med 2005 Sep 20;143(6):427-438. [doi: 10.7326/0003-4819-143-6-200509200-00007] [Medline: 16172441]
25. Murray E, Burns J, See T, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19;19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
26. Muller I, Rowsell A, Stuart B, Hayter V, Little P, Ganahl K, et al. Effects on engagement and health literacy outcomes of web-based materials promoting physical activity in people with diabetes: an international randomized trial. J Med Internet Res 2017 Jan 23;19(1):e21 [FREE Full text] [doi: 10.2196/jmir.6601] [Medline: 28115299]
27. Stellefson M, Chaney B, Chaney D. The digital divide in health education. Am J Health Educ 2013 Jan 23;39(2):106-112. [doi: 10.1080/19325037.2008.10599023]
28. Latulippe K, Hamel C, Giroux D. Social health inequalities and ehealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
29. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
30. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
31. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the health information national trends survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
32. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595. [doi: 10.1093/her/cyp063] [Medline: 19897515]
33. Hardiker NR, Grant MJ. Factors that influence public engagement with ehealth:a literature review. Int J Med Inform 2011 Jan;80(1):1-12. [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
34. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015 Jan 29;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
35. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cog Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
36. Gilmour JA. Reducing disparities in the access and use of internet health information. A discussion paper. Int J Nurs Stud 2007 Sep;44(7):1270-1278. [doi: 10.1016/j.ijnurstu.2006.05.007] [Medline: 16828775]
37. van Deursen A, van Dijk J. Internet skills and the digital divide. New Media Soc 2010 Dec 6;13(6):893-911. [doi: 10.1177/1461444810386774]
38. van Dijk J. The Deepening Divide: Inequality in the Information Society. Thousand Oaks, CA: Sage Publications; 2005.
39. Henderson L, Craig J, Willis N, Tovey D, Webster AC. How to write a Cochrane systematic review. Nephrology (Carlton) 2010 Sep;15(6):617-624. [doi: 10.1111/j.1440-1797.2010.01380.x] [Medline: 20883282]
40. Welch V, Petticrew M, Tugwell P, Moher D, O'Neill J, Waters E, PRISMA-Equity Bellagio Group. PRISMA-equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. PLoS Med 2012;9(10):e1001333 [FREE Full text] [doi: 10.1371/journal.pmed.1001333] [Medline: 23222917]
41. World Health Organization. The Global Burden of Disease: 2004 Update. Geneva, Switzerland: World Health Organization; 2004.

42. Goodwin N, Curry N, Naylor C, Ross S. The King's Fund: Ideas That Change Health and Care. 2010. Managing People With Long-term Conditions URL: https://www.kingsfund.org.uk/sites/default/files/field/field_document/managing-people-long-term-conditions-gp-inquiry-research-paper-mar11.pdf [accessed 2020-04-23]
43. Lewis R, Dixon J. Rethinking management of chronic diseases. Br Med J 2004 Jan 24;328(7433):220-222 [FREE Full text] [doi: 10.1136/bmj.328.7433.220] [Medline: 14739194]
44. Marmot M, Commission on Social Determinants of Health. Achieving health equity: from root causes to fair outcomes. Lancet 2007 Sep 29;370(9593):1153-1163. [doi: 10.1016/S0140-6736(07)61385-3] [Medline: 17905168]
45. Hoag CC, Elterman DS, Macneily AE. Abstracts presented at the American urological association annual meeting: determinants of subsequent peer reviewed publication. J Urol 2006 Dec;176(6 Pt 1):2624-9; discussion 2629. [doi: 10.1016/j.juro.2006.08.021] [Medline: 17085176]
46. Lucas PJ, Cabral C, Hay AD, Horwood J. A systematic review of parent and clinician views and perceptions that influence prescribing decisions in relation to acute childhood infections in primary care. Scand J Prim Health Care 2015 Mar;33(1):11-20 [FREE Full text] [doi: 10.3109/02813432.2015.1001942] [Medline: 25716427]
47. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [doi: 10.1016/0895-4356(90)90158-l] [Medline: 2348207]
48. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993 May;46(5):423-429. [doi: 10.1016/0895-4356(93)90018-v] [Medline: 8501467]
49. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. J Clin Epidemiol 2014 Jan;67(1):56-64. [doi: 10.1016/j.jclinepi.2013.08.005] [Medline: 24189091]
50. Kavanagh J, Oliver S, Lorenc T. ScienceOpen. 2008. Reflections on Developing and Using PROGRESS-Plus URL: https://www.scienceopen.com/document?vid=1492d18a-9f4d-4253-9ad9-9507848388a2 [accessed 2020-04-23]
51. Evans T, Brown H. Road traffic crashes: operationalizing equity in the context of health sector reform. Inj Control Saf Promot 2003;10(1-2):11-12. [doi: 10.1076/icsp.10.1.11.14117] [Medline: 12772480]
52. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
53. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
54. Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, et al. The harvest plot: a method for synthesising evidence about the differential effects of interventions. BMC Med Res Methodol 2008 Feb 25;8:8 [FREE Full text] [doi: 10.1186/1471-2288-8-8] [Medline: 18298827]
55. Kosse RC, Bouvy ML, de Vries TW, Koster ES. Effect of a mHealth intervention on adherence in adolescents with asthma: a randomized controlled trial. Respir Med 2019 Mar;149:45-51. [doi: 10.1016/j.rmed.2019.02.009] [Medline: 30803885]
56. Moy ML, Collins RJ, Martinez CH, Kadri R, Roman P, Holleman RG, et al. An internet-mediated pedometer-based program improves health-related quality-of-life domains and daily step counts in COPD: a randomized controlled trial. Chest 2015 Jul;148(1):128-137 [FREE Full text] [doi: 10.1378/chest.14-1466] [Medline: 25811395]
57. Voncken-Brewster V, Tange H, de Vries H, Nagykaldi Z, Winkens B, van der Weijden T. A randomized controlled trial evaluating the effectiveness of a web-based, computer-tailored self-management intervention for people with or at risk for COPD. Int J Chron Obstruct Pulmon Dis 2015;10:1061-1073 [FREE Full text] [doi: 10.2147/COPD.S81295] [Medline: 26089656]
58. Huang JS, Terrones L, Tompane T, Dillon L, Pian M, Gottschalk M, et al. Preparing adolescents with chronic disease for transition to adult care: a technology program. Pediatrics 2014 Jun;133(6):e1639-e1646 [FREE Full text] [doi: 10.1542/peds.2013-2830] [Medline: 24843066]
59. Joubert M, Armand C, Morera J, Tokayeva L, Guillaume A, Reznik Y. Impact of a serious videogame designed for flexible insulin therapy on the knowledge and behaviors of children with type 1 diabetes: the LUDIDIAB pilot study. Diabetes Technol Ther 2016 Feb;18(2):52-58. [doi: 10.1089/dia.2015.0227] [Medline: 26440963]
60. Pacaud D, Kelley H, Downey AM, Chiasson M. Successful delivery of diabetes self-care education and follow-up through ehealth media. Can J Diabetes 2012 Oct;36(5):257-262. [doi: 10.1016/j.jcjd.2012.08.006]
61. Raiff BR, Barrry VB, Ridenour TA, Jitnarin N. Internet-based incentives increase blood glucose testing with a non-adherent, diverse sample of teens with type 1 diabetes mellitus: a randomized controlled trial. Transl Behav Med 2016 Jun;6(2):179-188 [FREE Full text] [doi: 10.1007/s13142-016-0397-5] [Medline: 27198235]
62. Whittemore R, Jaser SS, Jeon S, Liberti L, Delamater A, Murphy K, et al. An internet coping skills training program for youth with type 1 diabetes: six-month outcomes. Nurs Res 2012;61(6):395-404 [FREE Full text] [doi: 10.1097/NNR.0b013e3182690a29] [Medline: 22960587]
63. Bahar-Fuchs A, Barendse M, Bloom R, Ravona-Springer R, Heymann A, Dabush H, et al. Computerized cognitive training for older adults at higher dementia risk due to diabetes: findings from a randomized controlled trial. J Gerontol A Biol Sci Med Sci 2020 Mar 9;75(4):747-754. [doi: 10.1093/gerona/glz073] [Medline: 30868154]

64. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]
65. Davis SA, Carpenter D, Cummings DM, Lee C, Blalock SJ, Scott JE, et al. Patient adoption of an internet based diabetes medication tool to improve adherence: a pilot study. Patient Educ Couns 2017 Jan;100(1):174-178. [doi: 10.1016/j.pec.2016.07.024] [Medline: 27480530]
66. Heinrich E, de Nooijer J, Schaper NC, Schoonus-Spit MH, Janssen MA, de Vries NK. Evaluation of the web-based diabetes interactive education programme (DIEP) for patients with type 2 diabetes. Patient Educ Couns 2012 Feb;86(2):172-178 [FREE Full text] [doi: 10.1016/j.pec.2011.04.032] [Medline: 21616626]
67. Lorig K, Ritter PL, Laurent DD, Plant K, Green M, Jernigan VB, et al. Online diabetes self-management program: a randomized study. Diabetes Care 2010 Jun;33(6):1275-1281 [FREE Full text] [doi: 10.2337/dc09-2153] [Medline: 20299481]
68. Yu CH, Parsons JA, Mamdani M, Lebovic G, Hall S, Newton D, et al. A web-based intervention to support self-management of patients with type 2 diabetes mellitus: effect on self-efficacy, self-care and diabetes distress. BMC Med Inform Decis Mak 2014 Dec 14;14:117 [FREE Full text] [doi: 10.1186/s12911-014-0117-3] [Medline: 25495847]
69. Istepanian R, Sungoor A, Earle KA. Technical and Compliance Considerations for Mobile Health Self-Monitoring of Glucose and Blood Pressure for Patients With Diabetes. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009 Presented at: IEMBS'09; September 3-9, 2009; Minneapolis, MN, USA. [doi: 10.1109/iembs.2009.5334580]
70. Offringa R, Sheng T, Parks L, Clements M, Kerr D, Greenfield MS. Digital diabetes management application improves glycemic outcomes in people with type 1 and type 2 diabetes. J Diabetes Sci Technol 2018 May;12(3):701-708 [FREE Full text] [doi: 10.1177/1932296817747291] [Medline: 29277103]
71. Nevedal DC, Wang C, Oberleitner L, Schwartz S, Williams AM. Effects of an individually tailored web-based chronic pain management program on pain severity, psychological health, and functioning. J Med Internet Res 2013 Sep 25;15(9):e201 [FREE Full text] [doi: 10.2196/jmir.2296] [Medline: 24067267]
72. Lawford BJ, Hinman RS, Kasza J, Nelligan R, Keefe F, Rini C, et al. Moderators of effects of internet-delivered exercise and pain coping skills training for people with knee osteoarthritis: exploratory analysis of the impact randomized controlled trial. J Med Internet Res 2018 May 9;20(5):e10021 [FREE Full text] [doi: 10.2196/10021] [Medline: 29743149]
73. Glasgow RE, Strycker LA, King DK, Toobert DJ. Understanding who benefits at each step in an internet-based diabetes self-management program: application of a recursive partitioning approach. Med Decis Making 2014 Feb;34(2):180-191. [doi: 10.1177/0272989X13498156] [Medline: 23913917]
74. Andrews T, Thompson M, Buckley DI, Heneghan C, Deyo R, Redmond N, et al. Interventions to influence consulting and antibiotic use for acute respiratory tract infections in children: a systematic review and meta-analysis. PLoS One 2012;7(1):e30334 [FREE Full text] [doi: 10.1371/journal.pone.0030334] [Medline: 22299036]
75. Vodicka TA, Thompson M, Lucas P, Heneghan C, Blair PS, Buckley DI, TARGET Programme Team. Reducing antibiotic prescribing for children with respiratory tract infections in primary care: a systematic review. Br J Gen Pract 2013 Jul;63(612):e445-e454 [FREE Full text] [doi: 10.3399/bjgp13X669167] [Medline: 23834881]
76. Thompson M, Vodicka T, Blair P, Buckley D, Heneghan C, Hay AD, TARGET Programme Team. Duration of symptoms of respiratory tract infections in children: systematic review. Br Med J 2013 Dec 11;347:f7027 [FREE Full text] [doi: 10.1136/bmj.f7027] [Medline: 24335668]
77. Cabral C, Horwood J, Hay AD, Lucas PJ. How communication affects prescription decisions in consultations for acute illness in children: a systematic review and meta-ethnography. BMC Fam Pract 2014 Apr 8;15:63 [FREE Full text] [doi: 10.1186/1471-2296-15-63] [Medline: 24708839]
78. Dwan K, Gamble C, Williamson PR, Kirkham JJ, Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias-an updated review. PLoS One 2013;8(7):e66844 [FREE Full text] [doi: 10.1371/journal.pone.0066844] [Medline: 23861749]
79. Appleby J, Deeming C. Data briefing. Inverse care law. Health Serv J 2001 Jun 21;111(5760):37. [Medline: 11436761]
80. Cartwright A, O'Brien M. Social class variations in health care and in the nature of general practitioner consultations. Sociol Rev Monogr 1976 Mar(22):77-98. [doi: 10.1111/j.1467-954x.1974.tb00019.x] [Medline: 1006355]
81. Buchan IC, Richardson IM. Time study of consultations in general practice. In: Scottish Health Services Statistics. Edinburgh, UK: Scottish Home and Health Department; 1973.
82. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
83. van Dijk J. The Evolution of the Digital Divide: the Digital Divide Turns to Inequality of Skills and Usage. Amsterdam, The Netherlands: IOS Press; 2012.
84. Rogers EM. Diffusion of Innovations. Fifth Edition. New York, USA: Free Press; 2003.
85. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality, and readability in English and Spanish. J Am Med Assoc 2001;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]

86. Collins PH, Blige SS. Intersectionality. Malden, MA: Polity Press; 2016.

Abbreviations

COPD: chronic obstructive pulmonary disease
NHS: National Health Service
NIHR: National Institute for Health Research
PP: PROGRESS Plus
QoL: quality of life
RCT: randomized controlled trial
RoB: risk of bias
ROBINS-I: Risk Of Bias In Non-randomized Studies-of Interventions
SDH: social determinants of health
SES: socioeconomic status
SPCR: School for Primary Care Research
T1D: type 1 diabetes
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 20.01.20; peer-reviewed by J Parsons, B Arnoldussen, E Sezgin; comments to author 02.03.20; revised version received 27.03.20; accepted 29.03.20; published 05.06.20.

Please cite as:
Turnbull S, Cabral C, Hay A, Lucas PJ
Health Equity in the Effectiveness of Web-Based Health Interventions for the Self-Care of People With Chronic Health Conditions: Systematic Review
J Med Internet Res 2020;22(6):e17849
URL: http://www.jmir.org/2020/6/e17849/
doi:10.2196/17849
PMID:32459632

Original Paper

Lifestyle Segmentation to Explain the Online Health Information–Seeking Behavior of Older Adults: Representative Telephone Survey

Winja Weber[1], MA; Anne Reinhardt[1], MA; Constanze Rossmann[1], PhD

[1] University of Erfurt, Erfurt, Germany

Corresponding Author:
Winja Weber, MA
University of Erfurt
Nordhaeuser Str 63
Erfurt
Germany
Phone: 49 361 737 4175
Email: winja.weber@uni-erfurt.de

Abstract

Background: As a result of demographic changes, the number of people aged 60 years and older has been increasing steadily. Therefore, older adults have become more important as a target group for health communication efforts. Various studies show that online health information sources have gained importance among younger adults, but we know little about the health-related internet use of senior citizens in general and in particular about the variables explaining their online health-related information–seeking behavior. Media use studies indicate that in addition to sociodemographic variables, lifestyle factors might play a role in this context.

Objective: The aim of this study was to examine older people's health-related internet use. Our study focused on the explanatory potential of lifestyle types over and above sociodemographic variables to predict older adults' internet use for health information.

Methods: A telephone survey was conducted with a random sample of German adults aged 60 years and older (n=701) that was quota-allocated by gender, age, educational status, and degree of urbanity of their place of residence.

Results: The results revealed that participants used the internet infrequently (mean 1.82 [SD 1.07]), and medical personnel (mean 2.89 [SD 1.11]), family and friends (mean 2.86 [SD 1.21]), and health brochures (mean 2.85 [SD 1.21]) were their main sources of health information. A hierarchical cluster analysis based on values, interests, and leisure time activities revealed three different lifestyle types for adults aged over 60 years: the Sociable Adventurer, the Average Family Person, and the Uninterested Inactive. After adding these types as second-step predictors in a hierarchical regression model with sociodemographic variables (step 1), the explained variance increased significantly (R^2=.02, P=.001), indicating that the Average Family Person and the Sociable Adventurer use the internet more often for health information than the Uninterested Inactive, over and above their sociodemographic attributes.

Conclusions: Our findings indicate that the internet still plays only a minor role in the health information–seeking behavior of older German adults. Nevertheless, there are subgroups including younger, more active, down-to-earth and family-oriented males that may be reached with online health information. Our findings suggest that lifestyle types should be taken into account when predicting health-related internet use behavior.

(J Med Internet Res 2020;22(6):e15099) doi:10.2196/15099

KEYWORDS

older adults; online health information seeking; lifestyle; segmentation; cluster analysis

Introduction

Background

The world population is aging; in 2017, at least 962 million people worldwide were 60 years old and older. Researchers expect a doubling of this number by 2050 [1]. The reasons for this tendency are an increase in life expectancy and a decline in birthrates. This demographic change implies multiple challenges for society, especially for the health care system and health care communicators because aging is linked with several health-related restrictions: research predicts a sharp increase in age-related diseases [2,3], which will lead to a greater need for medical care [4]. For this reason, it is important to effectively inform older adults about health issues, help them manage (chronic) diseases, and promote a healthy lifestyle and, in doing so, relieve the health care system in the long run. In order to guarantee successful communication, it is important to understand what sources of health information seniors use and which predictors best explain their health information–seeking behavior.

In general, the internet has become increasingly important as a source of health information. Research indicates that there has been a great increase in internet use for health information among several target groups [5-8]. Several studies even proclaim the internet to be the most important source of health information nowadays [7,8]. However, other studies show that sources such as interpersonal communication with health care providers or traditional media channels are still more important than the internet [9]. Among older adults, both data on general internet use [10] and data on online health information seeking [9,11] indicate that the internet does not play a major role as a source of health information among older adults.

Given the rapid evolution in this area and the contradictory view on the role of the internet for older adults, we sought to examine the role of the internet as a source of health information for older adults in Germany and, in particular, identify which attributes best explain their online health information–seeking behavior. Research findings indicate that sociodemographic variables are strong predictors of online health information behavior [11], but they do not explain all of it. Thus, explaining the use of the internet for health information using only sociodemographic variables reflects only a part of the picture. Studies in the context of consumer research have considered lifestyle attributes in order to better understand and explain user behavior [12]. Accordingly, online health information–seeking behavior might be influenced by lifestyle factors and values.

Thus, our study focused on the role of lifestyle variables to explain older adults' online health information behavior. In our article, we focus on adults aged 60 years and older because in the field of gerontology and consumer research it is a common practice to subdivide old age into those aged under 60 years and those aged 60 years or older [13-15]. Because there has been no research showing the impact of lifestyles as predictors of older adults' online health information behavior, we investigated in a representative survey (1) the role of the internet for health information–seeking behavior among German adults ages 60 years or older and (2) to what extent lifestyle types contribute to the model over and above sociodemographic variables.

Health Information–Seeking Behavior of Older Adults

According to the Health Information National Trends Survey, in 2014 at least 45% of Americans stated that the internet was their primary source of health information, followed by health care professionals, traditional media, and family and friends [7]. In Germany, representative data showed a similar picture—for instance, in 2013 40% of Germans used the internet for health information [16,17].

However, for adults aged 60 years and above, findings present a different picture. People in this age group predominantly consult health care professionals to obtain health-related information [18,19]. Moreover, they typically use traditional media such as print media, radio, TV, and material from health insurance funds or pharmacies for health-related information [17,20,21]. Although internet use for health-related information among older adults has been steadily increasing over the past years, it still seems to play a minor role. In 2013, just 16% of German adults 60 years and older indicated the use of the internet for health purposes [17]. In America, the number is slightly higher, with 29% of people 65 years and older using the internet to obtain health information [5]. Due to increasing internet use in general during the past several years and a lack of data about the current status quo, we asked:

- RQ1: Which sources of health information do older adults use in Germany and what role does the internet play in this context?

Factors Explaining Online Health Information–Seeking Behavior of Older Adults

Online health information seeking can be determined by many different factors. A representative study in Europe examined age as one of the strongest predictors of online health information behavior [11]. With increasing age, the number of people who use the internet to seek health-related information decreases [9,18]. Those differences between younger and older people might be a result of psychological and social aging processes. Thus, research found that older people have less technological control and self-efficacy in addition to fewer skills for adapting to new technologies [22]. This lack of confidence can reduce the intention to use online health information [23]. Additionally, adults do not trust the internet as a source of health information [9] and have difficulties in coping with the information overload they find when seeking information online [24]. Other age-related factors that determine internet use are people's cognitive capability (such as memory loss) and physical conditions (such as sight loss) [25]. Furthermore, people who are generally curious about life in particular search more actively for health information online [26].

In addition to age, research findings indicate that well-educated people who live in an urban area are more likely to use the internet for health-related purposes than those who are less educated and live in a rural area [11,16,27]. Research focusing on the influence of gender-specific online health information–seeking behavior demonstrates heterogeneous findings: while men generally use the internet more often, women are more interested in health information [28]. Accordingly, some researchers found more frequent health-related use of the internet for females [29,30], while others indicate that the typical online health information seeker is male [31].

Another discrepancy in research findings can be found with respect to individuals’ health status. On the one hand, studies indicate that people with poorer health are more likely to search for health information online than people with a good health status [11,26]. On the other hand, studies suggest that adults with a good health status use online health information sources more frequently than those with a poorer self-reported health status [32]. Given the inconclusive picture, we ask the following question:

- RQ2: Which factors (age, education, gender, place of residence, health status) explain older adults’ online health information–seeking behavior?

Lifestyle Segmentation to Explain Online Health Information–Seeking Behavior

Differences in (consumer) behavior cannot only be explained by demographic variables but rather are influenced by people’s general lifestyles as well [33]. Lifestyles can be defined as “patterns of action that differentiate people.... Lifestyles therefore help to make sense of...what people do, and why they do it, and what doing means to others” [34]. Lifestyle segmentation is a concept derived from market or audience research [12] that aims to cluster people with similar characteristics or values. It has been used primarily in the marketing sector [35] but also in the context of health campaign planning [12]. In marketing and consumer research literature, the term lifestyle is often used synonymously with psychographics [36]. However, psychographics mainly focus on personality traits whereas lifestyles mainly focus on needs and values closely associated with behavior [37]. Because of this, we decided to refer to the lifestyle concept instead of psychographics in this study.

In the context of lifestyles, audience segmentation can be seen as “classification schemes...[that] are based in overarching models and assumptions about...lifestyles and values and not on variables that predict a specific behavior” [38]. Lifestyle is primarily defined by activities, interests, and opinions [33,39-41]. Accordingly, activities (such as work or hobbies), interests (such as family, home, media), and beliefs (such as social issues, politics, culture) are typically measured in order to identify lifestyle types [42]. In addition, researchers often relate lifestyles to values [43] because they are a part of individuals’ views of the world and thus form a major factor in people’s character and behavior [33,44].

In the context of consumer research and audience segmentation, previous studies have already addressed lifestyle attributes to explain consumer behavior [33]. In the specific case of media use, lifestyle segmentation has been used to understand habits such as daily newspaper use [45], adolescents’ media use [46], adoption and use of pagers and mobile phones [41], older adults’ participation in online communities [47], and attitudes toward internet advertisements [48]. While these studies found differences in the behavior between the examined lifestyle types, another study investigated whether lifestyles have the potential to explain online news usage [49]. The authors found that lifestyles did not predict the adoption of online news in general. However, the enjoyment of interactive capability of online news was predicted most strongly by lifestyles in the group of adopters [49].

Most of the aforementioned studies included internet use either as a cluster-forming variable rather than a dependent variable [50,51] or studied differences within the clusters and not the explanatory potential of lifestyles [41,45-48]. Additionally, there is an ongoing debate about whether lifestyles, compared with demographics, are suitable to explain media use patterns. Skeptical researchers suggest that lifestyles are already included in demographic variables [52]. Hence, one study compared the influence of demographics and psychographics to explain the consumption of different forms of media use [53]. It found that the contribution of demographics and psychographics varied across the different behaviors: while psychographics played no role beyond the demographic explanation in outcomes such as genre preferences, the prediction of TV consumption benefited from adding lifestyle variables. Therefore, the authors of this study recommend integrating both approaches to explain and understand behavior [53].

The role of lifestyle variables becomes particularly important when focusing on older adults: several studies show that age itself is not the main predictor for behavior among older people [47,54,55]. A reason might be that aging is a very individualistic process: rather than aging chronologically, people age very differentially [54,56]. How old someone feels is directly associated with their mindset (eg, values) and activity—thus, lifestyle variables play a major role in explaining behavior [54]. For instance, a 65-year-old person can have a very inactive lifestyle and may wish to stay at home most of the time, while a 75-year-old person can be still very sporty and interested in discovering new things. This example might make clear why older people are often described as the most heterogeneous audience [47]. Lifestyle types have already been successfully used to segment this heterogeneous target group into homogenous subgroups to get a better understanding of the so-called gray market [50,51,57-59], but there is no research examining the impact of lifestyle segmentation on the online health information–seeking behavior older adults.

In sum, demographic variables such as age, education, gender, and health status have an influence on online health information–seeking behavior. However, in many cases lifestyle types can go beyond demographics in explaining behaviors. Nevertheless, there has been no research examining the impact of lifestyle segmentation on the online health information–seeking behavior in general and especially of older adults. This led us to ask the following:

- RQ3: Which lifestyle types can be identified in the age group 60 years and older based on general values, leisure values, leisure activities, and interests?
- RQ4: Do lifestyle types offer added value to explain older adults’ online health information–seeking behavior beyond demographics?

Methods

Recruitment

To answer our research questions, we conducted a telephone survey with 701 adults aged 60 years and older recruited by the Institute for Applied Marketing and Communication Research GmbH. As the study involved human participants, an institutional review board approval was obtained by the advisory board on ethical issues of the University of Erfurt (No. 17/05/29). All participants gave their informed consent to use and share their data for scientific purposes without disclosure of their identity. The respondents were selected randomly and quota-allocated by age (60-64 years, 65-74 years, and >75 years), rural/urban living areas (<10,000 residents, >10,000 residents), gender (male, female), and education (low, high). The survey was conducted as part of the formative research for the strategic development of a vaccination campaign and therefore also included health-related questions not relevant here. For this reason, the interview duration was about 30 minutes.

About half (15,967/31,419, 50.82%) of the total sample did not fit into the quota or were not reached. In addition, 46.57% (14,633/31,419) of the respondents refused to participate. Since the response rate was low (701/31,419 [2.23% of the gross sample]), the telephone survey did not fully reach the targeted quotas. Because of this, a weighting factor was applied in the following analyses, except in the factor analysis and cluster analysis—SPSS Statistics (IBM Corporation) does not allow the use of weighting in those data analytic procedures.

Measurement

Dependent Variables

Participants' health information–seeking behavior was measured using sources of health information with 11 items—internet use, health personnel, brochures, family and friends, conventional media (TV, radio, newspapers), apps, pharmacy, books, other patients, health insurance, and health workshops—on a 5-point scale (How often do you use the following sources of health information in general? 1 = never to 5 = very often) [60,61].

Lifestyle Variables

As a central determinant of lifestyle types, we measured general values and life goals with 11 items on a 5-point scale (1 = not at all important to 5 = very important): law and order, safety, success, creativity, political commitment, amusement, conformity, friends, environment, peace and harmony, and religion.

In addition, we assessed participants' leisure values with 3 items (experience adventures, meet interesting people, and do something new/crazy) on a 5-point scale (1 = not at all important to 5 = very important).

To examine the recreational activity of our participants, we asked how often they engage in the following leisure time activities during their free time (eg, at weekends, in the afternoon): using a computer, participating in sports, going to the theater/opera, visiting a museum, eating out in a restaurant, having guests visit them at home, reading books, watching TV, gardening, doing handicraft work, going for a walk, baking/cooking, listening to music, and spending time with their family. The frequency was measured on a 5-point scale (1 = never to 5 = very often).

Participants' personal interests were assessed on a 5-point scale (1 = not interested in to 5 = very interested in) via 11 items: politics, economics/law, sports, health, arts/literature, science, children/upbringing, partnership/family, traveling/vacation, work, and house/garden.

Further Predictor Variables

Participants' subjective health status was measured on a 5-point scale (1 = very bad to 5 = very good) with the question: How would you describe your actual state of health? (mean 3.06 [SD 0.80]) [62].

Finally, we assessed age, gender (0 = male, 1 = female), educational level (0 = low, including no educational qualifications, secondary modern school until class 9, or junior high; 1 = high, including university-entrance diploma, university degree, or higher), and the urbanity of place of residence (0 = under 10,000 residents, 1 = more than 10,000 residents).

Power Analysis

A power analysis determined the sample size for inferential statistics (independent *t* test: power=.80, α=.05, effect size *d*=0.2) and was rounded up to the nearest hundred [63].

Results

User Statistics

After weighting, the average age of participants was 71.6 (SD 7.3) years and 55.3% (388/701) were female. Most subjects had lower education levels (590/693, 86.0%) and came from cities with a population of more than 10,000 residents (518/701, 73.9%). Regarding their technical equipment, nearly two-thirds of our sample owned a computer (464/698, 66.2%) and nearly half of them had a smartphone (289/699, 41.4%). Furthermore, 60.7% of the participants (425/697) used the internet at least occasionally with an average duration of 85.2 (SD 109.9) minutes per day. See Multimedia Appendix 1 for descriptive analyses of all variables included in the following main analyses.

Evaluation Outcomes

Health Information Behavior of Older Adults

For our first research question (RQ1), we wanted to know what health information sources older adults use and if the internet plays a role in this context. A descriptive analysis (Table 1) showed that the most commonly used sources of health information were free brochures and pharmacy magazines, followed by interpersonal communication with medical staff, family, and friends, and traditional media sources. The internet was only infrequently used as a source of health information (mean 1.82 [SD 1.08]). Only other patients, health insurances, health workshops, and apps are used less often than the internet to get health information.

Table 1. Frequency of older adults' use of health information sources (n=681-689).

Source	Mean (SD)	95% CI
Health personnel	2.89 (0.04)	2.81-2.98
Family/friends	2.87 (0.40)	2.80-2.95
Brochures	2.85 (0.05)	2.76-2.94
Conventional media	2.70 (0.04)	2.61-2.78
Books	2.01 (0.04)	1.93-2.09
Pharmacy	1.97 (0.04)	1.90-2.04
Internet	1.83 (0.04)	1.78-1.91
Other patients	1.80 (0.04)	1.73-1.87
Health insurance	1.42 (0.03)	1.34-1.47
Health workshops	1.29 (0.03)	1.23-1.34
Apps	1.12 (0.02)	1.08-1.15

Lifestyle Types of Older Adults

In order to identify lifestyle types among adults aged 60 years and above (RQ3), we combined factor analyses with a cluster analysis. Because all lifestyle constructs were assessed via numerous items, in a first step we conducted exploratory factor analyses (rotation method: varimax; missing data: listwise deletion) with principal axis factoring in order to group similar variables into dimensions. Items that could not be summed up to one of the factors were treated separately in the cluster analysis.

Four of the 11 general values items could be combined into 2 factors (Kaiser-Meyer-Olkin [KMO]=.539; Multimedia Appendix 2), including the dimensions harmony (high loadings on importance of friends, peace/harmony) and regularity (high loadings on importance of law and order, safety).

The leisure activities were assessed with 14 items, of which 6 could be summed to 3 factors (KMO=.501; Multimedia Appendix 3), including culture (high loadings on going to a theater/opera, visiting museums), home and garden (high loadings on handicraft work, gardening), and technology (high loadings on using a computer, anti-baking/cooking). The leisure values (3 items) were treated separately (in accordance with the underlying scale).

The participants' personal interests were measured with 11 items that were combined into 2 factors (KMO=.580; Multimedia Appendix 4), including news (high loadings on interest in politics, economics/law) and family (high loadings on interest in children/upbringing, partnership/family, work/education).

In order to identify homogenous groups of cases based on the above-mentioned lifestyle variables, a hierarchical cluster analysis was computed. Ward minimum-variance clustering was performed using the squared Euclidean distance. Prior to the hierarchical cluster analysis, the relevant assumptions of this statistical analysis were tested [64,65]: (1) sample size of n=701 was deemed adequate, (2) statistical outliers were eliminated, (3) there were no missing data, (4) all analyzed variables were z-standardized, and (5) cluster variables were tested for normal distribution (the latter assumption was violated, but the hierarchical cluster analysis is robust to this violation) [66]. At each step, samples were merged into larger clusters to maximize the between-cluster sum of squares. Using this approach, three clusters were identified. They differed significantly in all cluster variables except for the leisure activity domestic work (Table 2).

Table 2. Differences in general values and life goals, leisure values, leisure activities, and interests between the identified clusters (n=595).

Variables	Total cohort, mean (SD)	Social Adventurer, mean (SD)	Average Family Person, mean (SD)	Uninterested Inactive, mean (SD)	F score	*P* value
General values						
Regularity	0.02 (0.79)	–0.19 (0.90)[a]	0.02 (0.77)[b]	0.30 (0.52)[c]	13.93	<.001
Harmony	0.01 (0.61)	0.07 (0.52)[a]	0.03 (0.62)[a]	–0.15 (0.66)[b]	5.05	.01
Success	0.00 (1.00)	0.34 (0.94)[a]	–0.14 (0.97)[b]	–0.14 (1.02)[b]	14.82	<.001
Creativity	0.01 (0.98)	0.45 (0.64)[a]	–0.02 (0.90)[b]	–0.54 (1.25)[c]	40.57	<.001
Political commitment	0.00 (0.99)	0.52 (0.94)[a]	–0.03 (0.97)[b]	–0.65 (0.65)[c]	57.78	<.001
Amusement	–0.01 (0.99)	0.56 (0.82)[a]	–0.11 (0.96)[b]	–0.56 (0.87)[c]	56.72	<.001
Conformity	–0.03 (0.96)	0.09 (0.97)[a]	–0.01 (1.01)[a]	–0.23 (0.77)[b]	3.87	.02
Environment	0.02 (0.99)	0.27 (0.78)[a]	–0.11 (1.10)[b]	–0.03 (0.91)[b]	8.32	<.001
Religion	–0.01 (1.00)	–0.00 (1.02)[a]	–0.01 (1.01)[a]	0.00 (0.92)[a]	0.01	>.99
Leisure values						
Experience adventures	–0.01 (0.99)	0.74 (1.00)[a]	–0.20 (0.88)[b]	–0.61 (0.50)[c]	100.48	<.001
Meet interesting people	0.02 (1.00)	0.51 (0.60)[a]	0.03 (0.94)[b]	–0.73 (1.15)[c]	64.23	<.001
Do something crazy	0.01 (0.99)	0.71 (1.01)[a]	–0.18 (0.87)[b]	–0.52 (0.69)[c]	82.91	<.001
Leisure activities						
Culture	0.01 (0.85)	0.59 (0.79)[a]	–0.03 (0.77)[b]	–0.70 (0.51)[c]	105.06	<.001
Technology	0.01 (0.63)	0.05 (0.66)[a]	0.03 (0.62)[a]	–0.12 (0.57)[b]	3.08	.047
Home and garden	0.03 (0.76)	0.10 (0.77)[a]	0.10 (0.75)[a]	–0.24 (0.72)[b]	10.01	<.001
Participating in sports	0.04 (0.99)	0.60 (0.76)[a]	0.03 (0.96)[b]	–0.75 (0.77)[c]	83.27	<.001
Eating out in a restaurant	0.01 (0.99)	0.28 (0.98)[a]	0.04 (0.95)[b]	–0.46 (0.95)[c]	20.44	<.001
Having guests	0.00 (0.98)	0.52 (0.75)[a]	–0.00 (0.90)[b]	–0.74 (1.01)[c]	70.30	<.001
Reading books	0.01 (1.00)	0.57 (0.71)[a]	0.09 (0.98)[b]	0.01 (0.90)[c]	51.87	<.001
Watching TV	0.01 (0.98)	–0.12 (1.02)[a]	0.03 (0.96)[a]	–0.75 (0.77)[a]	2.56	.08
Going for a walk	0.07 (0.96)	0.51 (0.74)[a]	0.00 (0.94)[b]	–0.39 (1.04)[c]	36.02	<.001
Listening to music	0.01 (0.99)	0.26 (0.88)[a]	0.07 (0.91)[a]	–0.52 (1.13)[b]	24.70	<.001
Spending time with family	0.03 (0.96)	0.28 (0.88)[a]	0.02 (0.96)[b]	–0.31 (0.97)[c]	13.62	<.001
Interests						
News	0.00 (0.82)	0.16 (0.73)[a]	0.08 (0.77)[a]	–0.46 (0.93)[b]	24.70	<.001
Family	0.01 (0.85)	0.27 (0.82)[a]	0.11 (0.76)[a]	–0.62 (0.81)[b]	48.21	<.001
Sports	0.01 (1.00)	0.39 (0.89)[a]	–0.07 (0.99)[b]	–0.32 (1.03)[c]	20.35	<.001
Health	0.02 (0.98)	0.18 (0.96)[a]	–0.03 (1.01)[b]	–0.09 (0.94)[b]	3.52	.03
Arts and literature	–0.02 (0.99)	0.55 (0.84)[a]	–0.04 (0.91)[b]	–0.80 (0.86)[c]	81.19	<.001
Science and technology	0.02 (1.00)	0.46 (0.81)[a]	–0.04 (0.99)[b]	–0.47 (1.04)[c]	32.59	<.001
Traveling and vacation	0.02 (1.00)	0.46 (0.81)[a]	0.09 (0.91)[b]	–0.82 (0.97)[c]	73.85	<.001
House and garden	0.02 (0.99)	–0.12 (1.01)[a]	0.20 (0.83)[b]	–0.24 (1.10)[a]	11.31	<.001

[a-c]Cells in a row with different letter superscripts differ with $P<.05$ (Duncan post hoc test).

Cluster 1 consisted of 28.5% (169/595) of the total subjects. This cluster was characterized by participants who were interested in almost all questioned activities. They stated an appreciation for meeting interesting people, doing something crazy, and experiencing adventures. In addition, they liked to have guests, read books, and spend time actively (eg, doing sports, visiting a museum). Furthermore, they attached great importance to a life full of creativity, amusement, and political engagement. Hence, cluster 1 was called the Sociable Adventurer.

The second cluster was the largest group, with 51.7% (308/595) of the total subjects. This group included older adults who put less value on experiencing adventures or doing something crazy. Instead, they appreciated doing domestic work and gardening or spending time with their families. In general, they did not focus on a life full of success or amusement. Therefore, we called this type the Average Family Person.

Cluster 3 was the smallest group (118/595, 19.8%) and significantly different from the others, characterized by participants who had no real interests and did not like to leave their comfort zones (such as having no interest in experiencing adventures or trying something new). These subjects had a great sense for regularity. Cluster 3 was called the Uninterested Inactive. For a more comprehensive overview of the three lifestyle types, we look at their demographics in the next step (Table 3).

Table 3. Characteristics of clusters by demographic variables.

Demographics	Sociable Adventurer, n (%)	Average Family Person, n (%)	Uninterested Inactive, n (%)
Age in years			
60-64	48 (28.4)	76 (24.8)	25 (21.2)
65-69	45 (26.6)	65 (21.2)	13 (11.0)
70-74	25 (14.8)	55 (17.9)	18 (15.3)
75+	51 (30.2)	111 (36.2)	62 (52.2)
Gender[a]			
Female	81 (47.9)	126 (41.0)	66 (55.9)
Male	88 (52.1)	181 (59.0)	52 (44.1)
Education[a]			
Low	67 (40.1)	172 (56.0)	87 (74.4)
High	100 (59.9)	135 (44.0)	30 (25.6)
Place of residence[a]			
Smaller city	23 (13.5)	94 (30.5)	39 (33.1)
Middle sized or larger city	147 (86.5)	214 (69.5)	79 (66.9)

[a] Demographic factors differ with $P<.05$ (χ^2 difference test).

Corresponding to the degree of activity, the oldest participants belonged to cluster 3. This cluster was also characterized by individuals with lower educational status. With regard to the place of residence, it becomes obvious that cluster 1 mainly consisted of people who lived in middle sized or larger cities. Last, cluster 2 can be characterized by a high proportion of men as compared with the other lifestyle types.

Factors Influencing Online Health Information–Seeking Behavior

In order to analyze which factors influence online health information seeking among older adults (RQ2 and RQ4), we conducted a 2-step hierarchical regression analysis with internet use for health information as the dependent variable (for an overview of distribution of the outcome variable, see Multimedia Appendices 1 and 5). The independent variables age, gender, educational status, health status, and place of residence were entered in step 1 and the lifestyle types in step 2. Because of the significantly different characterization of cluster 3, the Uninterested Inactive (dummy) served as reference group for the variables the Sociable Adventurer (dummy) and the Average Family Person (dummy). Prior to the hierarchical regression, the relevant assumptions of this statistical analysis were tested. First, the sample size was deemed adequate for the 7 independent variables to be included in the analysis [67]. The assumption of singularity was also met, as the independent variables were not a combination of other independent variables. An examination of correlations revealed that no independent variables were highly correlated. Also, there was no multicollinearity as indicated below (tolerance >0.10, variance inflation factor <10) [68]. Residual and scatter plots indicating the assumptions of normality, linearity, and homoscedasticity were all satisfied [69,70].

Hierarchical regression revealed that at step 1, demographic variables contributed significantly to the regression model (R^2=.114, $F_{5,582}$=14.92, P<.001). As can be seen in Table 3, age correlated negatively with the frequency of online health information seeking (β=–.236, P<.001). Accordingly, younger elderly used the internet more often than older ones as a source of health information. Another significant predictor was participant gender. Specifically, men showed greater online heath information–seeking behavior than woman (β=–.131, *P*=.001). Moreover, educational status (β=.126, *P*=.002) and health status (β=.081, *P*=.04) correlated positively and significantly with online health information use, indicating that highly educated and healthy seniors tend to use online health information more often than less well-educated subjects and those with poorer self-reported health. The place of residence was not significantly associated with health-related use of the internet.

Introducing the lifestyle types in step 2 explained another 2 percentage points of variance (R^2=.132). This change in R^2 was significant ($F_{7,580}$=12.57, *P*<.001). Again, the demographic variables age (β=–.208, *P*<.001), gender (β=–.151, *P*<.001), and education (β=.093, *P*=.03) correlated significantly with internet use, whereas the influence of subjects' health status disappeared. Both lifestyle types influenced online health information significantly; that is, the Sociable Adventurer (β=.186, *P*<.001) and the Average Family Person (β=.160, *P*=.003) used online health information more often than the Uninterested Inactive (Table 4). To examine differences between the Sociable Adventurer and the Average Family Person, we repeated the analysis with changed reference group (Multimedia Appendix 6). We found no significant influence of the Adventurer type when comparing it with the Average Family Person. Therefore, while both the Sociable Adventurer and the Average Family Person use the internet for health information more often than the Uninterested Inactive, there is no significant difference between them.

Table 4. Summary of hierarchical multiple regression analysis for variables predicting older adults' online health information–seeking behavior (n=587).

Predictors	Step 1				Step 2			
	B[a]	SE B[b]	β[c]	*P* value	B	SE B	β	*P* value
Step 1								
Age	–0.033	0.005	–0.236	<.001	–0.029	0.006	–0.208	<.001
Gender	–0.281	0.085	–0.131	.001	–0.323	0.085	–0.151	<.001
Education	0.270	0.087	0.126	.002	0.199	0.088	0.093	.03
Place of residence	0.184	0.096	0.076	.05	0.155	0.096	0.064	.11
Health status	0.102	0.050	0.081	.04	0.068	0.051	0.054	.18
Step 2								
Sociable Adventurer (dummy)	N/A[d]	N/A	N/A	N/A	0.439	0.132	0.186	.001
Average Family Person (dummy)	N/A	N/A	N/A	N/A	0.341	0.114	0.16	.003

[a] B: unstandardized regression coefficient.

[b] SE B: standard error for unstandardized regression coefficient.

[c] β: standardized regression coefficient.

[d] N/A: not applicable.

Due to the fact that internet use for health information did not show wide variance in our sample, we verified our results by testing a logistic regression model recoding internet use as a dummy variable (0 = never, 1 = at least infrequently). The tested model was significant (χ^2_2=9.853, *P*=.007). Findings indicate the same tendencies as the linear regression model (Multimedia Appendix 7).

Discussion

Principal Results and Comparison With Prior Work

Our study examined the relevance of the internet as a source of health information for older adults in Germany and the explanatory potential of lifestyle types over and above sociodemographic variables. Our findings highlight the fact that the internet still has a minor role compared with other sources in the health information–seeking behavior of older adults. Traditional health information sources such as brochures or medical consultations still are their major sources of health information. These findings are in line with current research showing that traditional media and health care workers are the most important health information sources for older adults [19,20].

A hierarchical cluster analysis based on values, interests, and leisure time activities revealed three different lifestyle types: (1) the first type was the Sociable Adventurer; these subjects were the most active with many interests and free time activities; (2) around 50% of the total participants were assigned to the Average Family Person; they preferred to do domestic work and gardening or spend time with their families; and (3) the smallest number of participants belonged to the Uninterested Inactive type; they were characterized by a lack of interests and an increased need for security.

There were several factors that abet the online information–seeking behavior of older adults. In line with current research, participants' age, education, and gender were, in particular, defining factors [11,27]: younger, highly educated participants used the internet more often for health information seeking than older participants with lower education. Regarding the heterogeneous evidence of the influencing role of gender on online health information–seeking behavior, our study suggests that the over 60-year-old male segment used the internet more often than women. This finding may be explained by the higher internet use of men in general [28] and in the age group of older adults in particular [71]. Furthermore, it seems reasonable that the age differences in internet use arose from cohort effects within the different age segments. A representative study showed that in 2017, Germans aged 50 to 69 years used the internet 98 minutes per day, while older adults (>70 years) used it much less (36 minutes per day) [72]. These numbers suggest that the internet will become an increasingly important health information source for the next generation of older people.

In addition, our study showed that beyond demographic variables, lifestyle types contributed significantly to our model and were relevant factors in predicting online health information–seeking behavior. Specifically, we found active and open-minded older adults (the Sociable Adventurer) as well as down-to-earth and family-oriented participants (the Average Family Person) to be more likely to seek online health information than uninterested, inactive ones (the Uninterested Inactive). This finding is in line with the results of another study indicating that people who are particularly curious and interested in many things use the internet more often for seeking health information [26]. Interestingly, by introducing the lifestyle types to the regression model, the influence of health status on the dependent variable disappeared. An explanation for this finding might be the fact that health status is associated with a person's lifestyle; lifestyle, in turn, includes many more dimensions (eg, interests, values) and seems to be more appropriate to predict the online health information–seeking behavior of older adults than health status alone.

The results indicate that demographics alone are not adequate to explain the use of internet sources for health information and lifestyles play an important role as well. Although the additional explained variance was rather small at 2%, lifestyle variables do explain significantly more than demographics alone. However, this small effect is in line with other studies that have either found no effect beyond sociodemographic variables (online news usage [49]) or only a very small effect (consuming various forms of media [53]). However, these studies also found that the explanatory potential of lifestyles varies between different behaviors [53] or may become more important if they are used to explain not only general media use but specific patterns within a behavior [49]. In addition, it should be noted that there are currently no other comparative studies, as most studies in this context rely on variance analyses to show differences in media use between different types of lifestyles [41,45-48]. Due to the ongoing debate about the suitability of lifestyles in comparison with demographics to explain media use patterns, our article provides new insight into this issue. Nevertheless, further studies are necessary to examine the influence of lifestyles beyond sociodemographic variables for different media use behaviors, especially among older people.

As a central part of health campaign planning, an adequate characterization of the addressees of a health message and a target group–specific implementation of the material often determines the success or failure of an intervention [73]. Our results showed that innovative digital health campaigns do not reach older adults as well as traditional media sources do. However, our study has shown that certain groups of older people can also be reached online; therefore, health information campaign planners should not only address sociodemographic variables but also general lifestyle types. Until now, only little was known about lifestyles in the age group 60 years and above except for a few studies such as one about preferences and perception of marketing communication of Poles aged 55 years and older [57] and a general lifestyle typology in the older American market [58]. Our study gives detailed insight into the lifestyles of older adults. This becomes relevant for health care communicators even outside older adults' online health information–seeking behavior. Therefore, addressing the Sociable Adventurer with images of active older people may work better than images of inactive testimonials. Another conclusion could be drawn for the implementation of health campaign material: in order to address the Average Family Person, gardening or living magazines could be used for distribution. These potential differences should be tested in another study.

Additionally, studies have shown that older people prefer a wide range of content when using the internet, depending on their lifestyle [74,75]. It can be assumed that not only the general internet use differs between lifestyle types but also the way the internet is used: Sociable Adventurers are the most active and open-minded among the examined lifestyle types, which is why they may be more likely to use blogs or interactive offerings. Many people aged 60 years and older, however, are more family oriented. Therefore, they may use the internet primarily for interpersonal communication with their loved ones. If so, social media or mailings have the potential to reach this specific group. We recommend that future studies should investigate the explanatory potential of lifestyle types for the use of specific health services in the internet rather than just general internet use for health information.

Limitations

This study has several limitations. First, we decided to focus on demographics and lifestyle types as defining factors of internet use for health information since other predictors such as trust, limited skills, or technophobia are defined as age-related [9,22]. However, the observed variables only explained 13% of the variance of older peoples' online health information–seeking behavior. Future research should include other variables such as trust in the internet or technical skills to better explain older adults' online health information–seeking behavior.

Further, the response rate was only 2.4%. Low response rates in telephone surveys can lead to biased data under the assumption that people who consistently participate in surveys are different than those who do not. This nonresponse bias occurs for civic and social engagement as well as volunteering [76]. Nevertheless, telephone polls show only minimal bias from nonresponses on lifestyle, health, and demographic questions, which leads us to the assumption that the nonresponse bias in our sample was low [76].

Last, the analyses are based on cross-sectional survey data. Therefore, we cannot make clear statements regarding causality and the sequence of relationships. Future research should focus on longitudinal data to investigate more support for the order of events.

Conclusions

Our study focused on understanding the use of the internet for health information by adults 60 years and older and how lifestyle types can contribute to explain this behavior. Our study showed that older adults still use the internet very infrequently when seeking health information. However, a subgroup of older adults may be reached with online health information; this group includes the subset of younger, male senior citizens who like to spend their free time actively or with their family and enjoy working in their house and garden. Taking into account a predicted increase in internet use by the next generation of older adults, these individuals should be adequate recipients of online health information, while older inactive people should still be addressed with traditional media like health information brochures.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) under grant number 03ZZ0819A, the Robert Koch Institute, the University Hospital of Jena, and Lindgrün GmbH. The total volume of the grant is 2,136,463 €. The Federal Ministry for Education and Research did not influence the design of the study; the collection, analysis, and interpretation of data; or the writing of the manuscript. The authors would like to thank the members of the vaccination60+ study group and its advisory board.

Authors' Contributions

All authors contributed to the conceptualization of the study and the writing and review of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive analysis of all variables considered in the analysis.
[DOCX File , 17 KB - jmir_v22i6e15099_app1.docx]

Multimedia Appendix 2
Factor analysis of general values constructs.
[DOCX File , 13 KB - jmir_v22i6e15099_app2.docx]

Multimedia Appendix 3
Factor analysis of leisure activity constructs.
[DOCX File , 14 KB - jmir_v22i6e15099_app3.docx]

Multimedia Appendix 4
Factor analysis of interest constructs.
[DOCX File , 13 KB - jmir_v22i6e15099_app4.docx]

Multimedia Appendix 5
Descriptive analysis of the outcome variable for the three clusters.
[DOCX File , 13 KB - jmir_v22i6e15099_app5.docx]

Multimedia Appendix 6
Summary of hierarchical multiple regression analysis for variables predicting older adults’ online health information seeking behavior: changed reference group.
[DOCX File , 14 KB - jmir_v22i6e15099_app6.docx]

Multimedia Appendix 7
Summary of logistic regression analysis for variables predicting older adults’ online health information seeking behavior.
[DOCX File , 14 KB - jmir_v22i6e15099_app7.docx]

References

1. World population ageing 2017. New York: United Nations Department of Economic and Social Affairs, Population Division URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf [accessed 2020-05-20]
2. Bleibler F, Konnopka A, Benzinger P, Rapp K, König H. The health burden and costs of incident fractures attributable to osteoporosis from 2010 to 2050 in Germany—a demographic simulation model. Osteoporos Int 2013 Mar;24(3):835-847. [doi: 10.1007/s00198-012-2020-z] [Medline: 22797490]
3. Peters E, Pritzkuleit R, Beske F, Katalinic A. [Demographic change and disease rates: a projection until 2050]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2010 May;53(5):417-426. [doi: 10.1007/s00103-010-1050-y] [Medline: 20376420]
4. Siewert U, Fendrich K, Doblhammer-Reiter G, Scholz RD, Schuff-Werner P, Hoffmann W. Health care consequences of demographic changes in Mecklenburg-West Pomerania: projected case numbers for age-related diseases up to the year 2020, based on the Study of Health in Pomerania (SHIP). Dtsch Arztebl Int 2010 May;107(18):328-334 [FREE Full text] [doi: 10.3238/arztebl.2010.0328] [Medline: 20517531]
5. Fox S. The Social Life of Health Information, 2011. Washington: Pew Internet and American Life Project; 2011 May 12. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf [accessed 2020-05-20]
6. Fox S, Duggan M. Health Online 2013. Washington: Pew Internet and American Life Project; 2013 Jan 15. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2019-04-30]
7. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
8. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
9. Chaudhuri S, White C, Thompson H, Demiris G. Examining health information-seeking behaviors of older adults. Comput Inform Nurs 2013 Nov;31(11):547-553 [FREE Full text] [doi: 10.1097/01.NCN.0000432131.92020.42] [Medline: 23974574]
10. Internet/Broadband Fact Sheet. Washington: Pew Internet and American Life Project; 2019 Jun 12. URL: https://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2019-04-30]
11. Bachl M. Online health information seeking in Europe: do digital divides persist? SCM Stud Commun Media 2016;5(4):427-453. [doi: 10.5771/2192-4007-2016-4-427]
12. Grunig J. Publics, audiences, and market segments: segmentation principles for campaigns. In: Salmon C, editor. Information Campaigns: Balancing Social Values and Social Change. Newbury Park: Sage; 1989:199-228.
13. Kroeber-Riel W, Weinberg P, Gröppel-Klein A. Konsumentenverhalten [Consumer Behavior]. München: Verlag Franz Vahlen; 2009.
14. Martin M, Kliegel M. Psychologische Grundlagen der Gerontologie [Psychological Basics of Gerontology]. Stuttgart: Kohlhammer; 2014.
15. Meyer-Hentschel H, Meyer-Hentschel G. Seniorenmarketing: Generationsgerechte Entwicklung und Vermarktung von Produkten und Dienstleistungen [Senior marketing: Generational development and marketing of products and services]. Göttingen: Business Village; 2009.
16. Baumann E, Czerwinski F. Erst mal Doktor Google fragen? Nutzung neuer Medien zur Information und zum Austausch über Gesundheitsthemen [First ask Doctor Google? Use of new media for information and exchange on health topics]. In: Böcken J, Braun B, Meierjürgen R, editors. Gesundheitsmonitor 2015. Bürgerorientierung im Gesundheitswesen [Health Monitor 2015: Citizen Orientation in Health Care]. Gütersloh: Verlag Bertelsmann Stiftung; 2015:57-79.
17. mindline media GmbH. Informationsquellen Zum Thema “Gesundheit Und Medizin.” Im Trendvergleich [Sources of information on “health and medicine.” A comparison of trends] URL: http://mindline-media.de/fileadmin/presse_downloads/Chartbericht_Infoquellen_Gesundheit_2013.pdf [accessed 2019-05-03]
18. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
19. Hall AK, Bernhardt JM, Dodd V. Older adults' use of online and offline sources of health information and constructs of reliance and self-efficacy for medical decision making. J Health Commun 2015;20(7):751-758 [FREE Full text] [doi: 10.1080/10810730.2015.1018603] [Medline: 26054777]
20. Massey PM. Where do U.S. adults who do not use the internet get health information? Examining digital health information disparities from 2008 to 2013. J Health Commun 2016;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]
21. Roski R, Schikorra S. Health Literacy: Wie informieren sich alte Menschen über Gesundheitsthemen? [Health Literacy: How do old people inform themselves about health issues?]. Public Health Forum 2007;15(4):20-23. [doi: 10.1016/j.phf.2007.10.012]
22. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]

23. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
24. Manafo E, Wong S. Exploring older adults' health information seeking behaviors. J Nutr Educ Behav 2012;44(1):85-89. [doi: 10.1016/j.jneb.2011.05.018] [Medline: 22101129]
25. Lorenz A, Oppermann R. Mobile health monitoring for the elderly: designing for diversity. Pervasive Mob Comput 2009 Oct;5(5):478-495. [doi: 10.1016/j.pmcj.2008.09.010]
26. Pang P, Chang S, Pearce J, Verspoor K. Online health information seeking behaviour: understanding different search approaches. Pac Asia Conf Inf Syst 2014:24-28.
27. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
28. Hallyburton A, Evarts LA. Gender and online health information seeking: a five survey meta-analysis. J Consum Health Internet 2014 Jun 04;18(2):128-142. [doi: 10.1080/15398285.2014.902268]
29. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
30. Smith D. Health care consumer's use and trust of health information sources. J Commun Healthc 2011 Oct;4(3):200-210. [doi: 10.1179/1753807611Y.0000000010]
31. Kim S. An exploratory study of inactive health information seekers. Int J Med Inform 2015 Feb;84(2):119-133 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.003] [Medline: 25453277]
32. Ghweeba M, Lindenmeyer A, Shishi S, Abbas M, Waheed A, Amer S. What predicts online health information-seeking behavior among Egyptian adults? a cross-sectional study. J Med Internet Res 2017 Jun 22;19(6):e216 [FREE Full text] [doi: 10.2196/jmir.6855] [Medline: 28642216]
33. Vyncke P. Lifestyle Segmentation. Eur J Commun 2002;17(4):445-463. [doi: 10.1177/02673231020170040301]
34. Chaney D. Lifestyles. New York: Routledge; 1996.
35. Smith W. Product differentiation and market segmentation as alternative marketing strategies. In: Walters C, Robin D, editors. Classics in Marketing. Santa Monica: Goodyear; 1979.
36. Lawson R, Todd S. Consumer lifestyles: a social stratification perspective. Marketing Theory 2016 Aug 11;2(3):295-307. [doi: 10.1177/1470593102002003278]
37. Wells W. Life Style and Psychographics. Chicago: American Marketing Association; 2011.
38. Slater MD. Theory and method in health audience segmentation. J Health Commun 1996;1(3):267-283. [doi: 10.1080/108107396128059] [Medline: 10947364]
39. González AM, Bello L. The construct “lifestyle” in market segmentation. Eur J Mark 2002 Feb;36(1/2):51-85. [doi: 10.1108/03090560210412700]
40. Peter J, Olson J. Understanding Consumer Behavior. Irwin: Burr Ridge; 1994.
41. Wei R. Lifestyles and new media: adoption and use of wireless communication technologies in China. New Media Soc 2016 Jun 30;8(6):991-1008. [doi: 10.1177/1461444806069879]
42. Otte G. Sozialstrukturanalysen mit Lebensstilen. Eine Studie zur Theoretischen und methodischen Neuorientierung der Lebensstilforschung [Social structure analyses with lifestyles. A study on the theoretical and methodical reorientation of lifestyle research]. Wiesbaden: VS Verlag für Sozialwissenschaften; 2004.
43. Mitchell A. The Nine American Lifestyles: Who We Are and Where We're Going. New York: Warner Books; 1984.
44. Solomon M. Consumer Behavior. Buying, Having, Being. London: Prentice Hall; 2017.
45. Schwartz SH, Moore RL, Krekel TH. Life style and the daily paper: a psychographic profile of midwestern readers. Newsp Res J 2016 Aug;1(1):9-18. [doi: 10.1177/073953297900100102]
46. Mayer C, Meinzolt A, Rossmann C, Brosius H. Der Einfluß von Lebenswelten Jugendlicher auf ihre Mediennutzung [The influence of adolescents' lifeworlds on their media use behavior]. In: Hasebrink U, Rössler P, editors. Publikumsbindungen. Medienrezeption zwischen Individualisierung und Integration [Audience Retention. Media Reception between Individualization and Integration]. München: Fischer; 1999:147-166.
47. Nimrod G. Applying gerontographics in the study of older internet users. J Aud Recept Stud 2013;10(2):19 [FREE Full text]
48. Yang KCC. A comparison of attitudes towards Internet advertising among lifestyle segments in Taiwan. J Mark Commun 2004 Sep;10(3):195-212. [doi: 10.1080/1352726042000181657]
49. Chan JK, Leung L. Lifestyles, reliance on traditional news media and online news adoption. New Media Soc 2016 Jun 30;7(3):357-382. [doi: 10.1177/1461444805052281]
50. DIVSI Ü60-Studie. 2016. Die digitalen Lebenswelten der über 60-Jährigen in Deutschland URL: https://www.divsi.de/wp-content/uploads/2016/10/DIVSI-UE60-Studie.pdf [accessed 2019-04-30]
51. Vuori S, Holmlund‐Rytkönen M. 55+ people as internet users. Mark Intell Plan 2005 Jan;23(1):58-76. [doi: 10.1108/02634500510577474]

52. Haas A. Medienmenüs. Der Zusammenhang zwischen Mediennutzung, SINUS-Milieus und Soziodemographie. München: Verlag Reinhard Fischer; 2007.
53. Sandy CJ, Gosling SD, Durant J. Predicting consumer behavior and media preferences: the comparative validity of personality traits and demographic variables. Psychol Mark 2013 Oct 06;30(11):937-949. [doi: 10.1002/mar.20657]
54. Moschis GP. Marketing to older adults: an updated overview of present knowledge and practice. J Consum Mark 2003 Nov;20(6):516-525. [doi: 10.1108/07363760310499093]
55. Moschis GP, Lee E, Mathur A. Targeting the mature market: opportunities and challenges. J Consum Market 1997 Aug;14(4):282-293. [doi: 10.1108/07363769710188536]
56. Filipp S, Mayer A. Bilder des Alters. Altersstereotype und die Beziehungen zwischen den Generationen [Pictures of Aging. Age Stereotypes and the Relationships Between Generations]. Stuttgart: Kohlhammer; 1999.
57. Maison D. Values and lifestyle segmentation of Poles 55+ and its implications for consumers' attitudes and behaviour. Manag Bus Adm Cent Eur 2014 Dec 15;22(4):78-94. [doi: 10.7206/mba.ce.2084-3356.121]
58. Sorce P, Tyler PR, Loomis LM. Lifestyles of older Americans. J Consum Mark 1989 Mar;6(3):53-63. [doi: 10.1108/EUM0000000002555]
59. Sudbury L, Simcock P. A multivariate segmentation model of senior consumers. J Consum Mark 2009 Jun 26;26(4):251-262. [doi: 10.1108/07363760910965855]
60. Gesundheitsberichterstattung des Bundes. Nutzung von Informationsquellen über Gesundheit. URL: http://www.gbe-bund.de/gbe10/abrechnung.prc_abr_test_logon?p_uid=gast&p_aid=0&p_knoten=FID&p_sprache=D&p_suchstring=9778 [accessed 2019-05-02]
61. Lenzner T, Neuert C, Hadler P. Krankheitswissen und Informationsbedarfe—Diabetes Mellitus: Fragebogen für Personen mit Diabetes; Kognitiver Pretest [Disease Knowledge and Information Needs—Diabetes Mellitus: Questionnaire for Persons with Diabetes: Cognitive Pretest]. Mannheim: GESIS—Leibniz-Institut für Sozialwissenschaften; 2017.
62. Garcia P, McCarthy M. Measuring Health. A Step in the Development of City Health Profiles. Copenhagen: World Health Organization; 1996.
63. Betsch C, Rossmann C, Pletz MW, Vollmar HC, Freytag A, Wichmann O, et al. Increasing influenza and pneumococcal vaccine uptake in the elderly: study protocol for the multi-methods prospective intervention study Vaccination60+. BMC Public Health 2018 Jul 16;18(1):885 [FREE Full text] [doi: 10.1186/s12889-018-5787-9] [Medline: 30012141]
64. Cleff T. Deskriptive Statistik und Explorative Datenanalyse [Descriptive Statistics and Explorative Data Analysis]. Wiesbaden: Gabler; 2015.
65. Schendera C. Clusteranalyse mit SPSS: Mit Faktorenanalyse [Cluster analysis using SPSS: including factor analysis]. München: Oldenbourg Verlag; 2010.
66. Norusis M. IBM SPSS Statistics 19 Statistical Procedures Companion. Upper Saddle River: Prentice Hall; 2012.
67. Tabachnick B, Fidell L. Using Multivariate Statistics. New York: Pearson; 2013.
68. Bühner M, Ziegler M. Statistik Für Psychologen Und Sozialwissenschaftler [Statistics for Psychologists and Social Scientists]. München: Pearson; 2009.
69. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis. A Global Perspective. New York: Pearson; 1998.
70. Pallant J. SPSS Survival Manual. Buckingham: Open University Press; 2001.
71. van Deursen AJ, Helsper EJ. A nuanced understanding of Internet use and non-use among the elderly. Eur J Commun 2015 Apr;30(2):171-187. [doi: 10.1177/0267323115578059]
72. Koch W, Frees B. ARD/ZDF-Onlinestudie 2017: Neun von zehn Deutschen online [ARD/ZDF online study: Nine out of ten Germans are onliners]. Media Perspekt 2017;9:434-446 [FREE Full text] [doi: 10.1007/978-3-531-93043-5_17]
73. Silk K, Atkin C, Salmon C. Developing effective media campaigns for health promotion. In: Thompson T, Parrott R, Nussbaum J, editors. The Routledge Handbook of Health Communication. London: Routledge; 2011:203-219.
74. van Boekel LC, Peek ST, Luijkx KG. Diversity in older adults' use of the internet: identifying subgroups through latent class analysis. J Med Internet Res 2017 May 24;19(5):e180 [FREE Full text] [doi: 10.2196/jmir.6853] [Medline: 28539302]
75. Gatto SL, Tak SH. Computer, internet, and e-mail use among older adults: benefits and barriers. Educ Gerontol 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
76. Keeter S, Hatley N, Kennedy C. What low response rates mean for telephone surveys. Washington: Pew Internet and American Life Project; 2017. URL: http://assets.pewresearch.org/wp-content/uploads/sites/12/2017/05/12154630/RDD-Non-response-Full-Report.pdf [accessed 2019-04-30]

Abbreviations

KMO: Kaiser-Meyer-Olkin

Edited by G Eysenbach; submitted 19.06.19; peer-reviewed by AL Camerini, PCI Pang; comments to author 03.10.19; revised version received 28.11.19; accepted 12.04.20; published 12.06.20.

Please cite as:
Weber W, Reinhardt A, Rossmann C
Lifestyle Segmentation to Explain the Online Health Information–Seeking Behavior of Older Adults: Representative Telephone Survey
J Med Internet Res 2020;22(6):e15099
URL: http://www.jmir.org/2020/6/e15099/
doi:10.2196/15099
PMID:32530433

Original Paper

Views of Swedish Elder Care Personnel on Ongoing Digital Transformation: Cross-Sectional Study

Katarina Baudin[1], MSc; Christine Gustafsson[1], PhD; Susanne Frennert[2], PhD

[1]School of Health, Care and Social Welfare, Mälardalen University, Eskilstuna, Sweden

[2]Internet of Things and People Research Center, Department of Computer Science and Media Technology, Malmö University, Malmö, Sweden

Corresponding Author:
Katarina Baudin, MSc
School of Health, Care and Social Welfare
Mälardalen University
Box 325
Eskilstuna, SE-631 05
Sweden
Phone: 46 735029494
Email: katarina.baudin@mdh.se

Abstract

Background: Swedish municipalities are facing demographic challenges due to the growing number of older people and the resulting increased need for health care services. Welfare technologies are being launched as possible solutions for meeting some of these challenges.

Objective: The aim of this study was to explore the perception, experimentation, evaluation, and procurement of welfare technology practices among professionals working in municipal elder care in relation to their gender, age, and profession.

Methods: Data for this explorative cross-sectional study were collected from 393 responses to a web-based survey on municipal elder care in Sweden. Chi square tests were performed to determine the associations.

Results: The results revealed gender, age, and professional differences in perspectives of municipal elder care workers. Differences were particularly evident in attitudes toward technology, both the use of technology in general and in the workplace, and involvement and participation in decision making regarding the procurement of new welfare technologies. Men (37/53, 70%) expressed a more positive attitude toward and curiosity regarding new technologies than women (157/336, 46.7%) (*P*=.03). Regarding age, the younger respondents (18-24 years old) perceived the digital transformation in the workplace as “too slow” (4/4, 100%), whereas the majority of older respondents (65-74 years old) perceived it as happening at the “right pace” (4/7, 57%). The elder care personnel felt encouraged by management to explore and experiment with new welfare technologies, but never did so either for management or with patients. Even though the majority of the respondents were women, more men (4/7, 57%) were involved in the procurement process for welfare technology devices and solutions than women (98/336, 29.2%) (*P*<.001).

Conclusions: Personnel working within municipal elder care were generally very positive toward new technologies. However, both gender and age differences may influence these perspectives such as the personnel’s resistance to welfare technology and patients’ participation in welfare technology usage and deployment. Different levels of participation in the decision-making process regarding new technology deployment may negatively affect the overall digital transformation within municipal elder care.

(J Med Internet Res 2020;22(6):e15450) doi:10.2196/15450

KEYWORDS

elder care; welfare technology; gender; municipality; participation; digitalization; age

Introduction

Background

Due to demographic changes, developed countries have growing concerns about the future challenges aging populations will present to their welfare systems [1-3]. In the Nordic countries, welfare technologies have been introduced as an important means of meeting these challenges [1]. In Sweden, the welfare system was developed to promote universal rights and social equality. Despite the increasing demand on the limited resources of welfare services, it is believed that welfare technology can help to sustain these rights through the digital transformation of care for all Swedish citizens, including older people in need [4]. The Swedish National Board of Health and Welfare believes that digital technology can help older individuals and people with disabilities to feel safe and to participate in society [5].

The expectation is that welfare technology will "gain time" for health care personnel to engage in human contact and that it will increase patients' self-management and independence [6-8]. Societal changes and scarce resources are constant challenges to both welfare technology practitioners and policy makers whose aim is to reap the benefits of technological change for individuals and for society as a whole [9-12]. Municipalities and county councils or regions in Sweden provide welfare technology as aids and housing adaptations. Examples of welfare technologies provided at the municipal level are social security alarms that can be used to call for emergency help, electronic home services that replace or supplement physical visits with digital contacts, advanced toilets with flush and drying functions, key-free home services that replace physical keys with digital key management systems, individual rehabilitation training through game consoles in the home, camera surveillance during the night, and medicine reminders of when to take prescriptions, which may be linked to an alarm system [13]. Welfare technology designed to support daily living must be understood and assessed in their context and from a social perspective [1,13,14]. The provision of a good lifestyle through welfare technology is affected by decisions made at the municipal level; it is therefore important to understand how the decisions are made at the municipal level when new welfare technologies are purchased.

Exploring the role of welfare technology in municipal elder care is important because the use of technology in home care and in health care is constantly increasing [12,14-21]. However, research on the use of welfare technology in elder care has often focused on the attributes of leadership and management. A review of welfare technology use in elder care highlighted that the success factors in integrating welfare technology into elder care are having clear goals, incentives, and strong leadership; an infrastructure with proper capacity; a well-functioning organizational structure; collaboration with others; and awareness of resources [14]. However, the review did not identify success factors relating to gender, age, or work experience among the various care professionals involved in welfare technology work and in decision making regarding welfare technology.

Aim

The aim of this study was to explore the perceptions of professionals involved in municipal elder care regarding the use of welfare technology and to assess the influence of gender, age, and profession on their perceptions, experimentation, evaluation, and procurement of these technologies.

Methods

Survey Design

This explorative cross-sectional study involved a web survey and a descriptive data analysis. In May 2018, the authors sent a hyperlink to a web-based survey to all municipal registrars in Sweden (N=290). The registrars were asked to distribute the link to those working with and involved in welfare technology work in elder care organizations. In the information letter, the researchers stated that that the respondents' answers would be confidential and not linked to any individual. Participation was voluntary. Anonymity was assured and participants were informed of the survey aim.

To explore the aim of the study, in the absence of any previous research addressing this topic, a new questionnaire was designed, and its psychometric properties were tested for both validity and reliability [22]. The survey covered five question areas with a fixed number of options (3-6 possible answers) for the respondent to choose from. The areas included questions on perceptions of the speed of technological change, participation in decision making in relation to welfare technology, experimentation and exploration of welfare technology at work, involvement in procurement, and "about you." The survey also included four areas of open-ended questions addressing the concept, advantages and potential, barriers to use, and evaluation methods for welfare technology. The results of the open-ended questions have been published in a separate article [23]. The questionnaire was pilot-tested with three potential users, and its content, face validity, and test-retest reliability were examined prior to distribution [24]. Four established and experienced researchers rated the content validity of the items using a 4-point scale (4=very relevant, 3=quite relevant, 2=somewhat relevant, 1=not relevant). The same experts were asked to determine the face validity by answering the following question: "Please write your reflections and comments about the questionnaire concerning readability, clarity, and layout." After 3 weeks, their comments on the content and face validity were reviewed by the authors and discussed [24], and some changes were made to the questionnaire. The test-retest reliability tests, conducted at 2-week intervals with three potential users, showed that the questionnaire was easy to understand and answer and took 10-15 minutes to complete [25]. The focus of the present study is to analyze the quantitative data from the fixed-option questions (see Multimedia Appendix 1) in the questionnaire.

Sample

In this explorative cross-sectional study [22], the respondents were personnel who worked with welfare technology in municipal elder care. Table 1 summarizes the respondents' professional affiliations.

Table 1. Characteristics of the study population (N=393).

Characteristic	n (%)
Sex	
Male	53 (13.5)
Female	336 (85.5)
Missing	4 (1.0)
Age (years)	
18-24	4 (1.0)
25-34	43 (10.9)
35-44	75 (19.1)
45-54	146 (37.2)
55-64	118 (30.0)
65-74	7 (1.8)
Working experience	
Less than 1 year	18 (4.6)
1-4 years	52 (13.2)
5-9 years	58 (14.8)
10-19 years	125 (31.8)
More than 20 years	140 (35.6)
Profession/Roles	
Information technology staff	29 (7.4)
Chief medical nurse responsible	78 (19.8)
Chief rehabilitation office responsible	28 (7.1)
Occupational therapist/physiotherapist	103 (26.2)
Specialist dementia nurse	51 (13.0)
Other (manager/electronic health strategist)	104 (26.5)

Statistical Analysis

Descriptive statistics were used to summarize the characteristics of the sample and its subgroups. Crosstabulation with chi square computation was used for analysis of categorical variables, which is a nonparametric test recommended when analyzing ordered categorical data [22].

The associations between the variables are summarized in Multimedia Appendix 2. The occupational therapist (OT) and physiotherapist (PT) respondents who have direct contact with patients in their everyday work were grouped into the category "OT/PT." Managers and electronic health (eHealth) strategists who work at the strategic level in elder care organizations and do not have daily contact with patients were grouped as "other." The survey data were imported into and analyzed with IBM SPSS version 24.

Ethics

This study did not include any personal or sensitive information that required ethical approval under the standards of the Swedish Research Council [26]. The study followed the guidelines for research ethics issued by the Swedish Research Council [26].

Results

Respondents' Characteristics

The online survey elicited 393 responses. Table 1 presents the respondents' demographic data.

Perceptions of the Speed of Change

The perceptions of personnel in elder care regarding the speed of digital transformation in their organizations was investigated through a question asking the participants to "grade the speed of the digital transformation in your workplace" (see questionnaire, Multimedia Appendix 1). The majority of both women (245/336, 72.9%) and men (46/53, 87%) answered that it was "too slow," and the chi square test showed a significant association of gender (*P*=.05). Regarding age, all of the younger respondents (18-24 years old) perceived the speed of digital transformation as "too slow" (4/4, 100%), whereas the majority of older respondents (65-74 years) perceived the digital transformation in the workplace as happening at the "right pace" (4/7, 57%).

The perceptions of surveyed personnel regarding the overall speed of technological change in health care was also examined

(Multimedia Appendix 1: question 2), demonstrating an association between gender and attitudes toward technology and digitalization within health care organizations (*P*=.03). Although 157/336 (46.7%) of the women disagreed with the statement that the change in health care organizations was too rapid, 37/53 (70%) of the men also disagreed with the statement. The results also showed an association between the respondents' profession/role in the organization and their perception as to whether technology and digitalization have made health care organizations change too quickly (*P*=.04). Specifically, 21/29 (72%) of the information technology (IT) staff, 20/28 (71%) of the chief rehabilitation officers responsible, and 42/78 (54%) of the chief medical nurses responsible disagreed with the statement that technology and digitalization have made health care organizations change too quickly. These results indicate that in their experience, technological developments in elder care have not been occurring fast enough.

The questionnaire also addressed whether the personnel perceived their workplaces as using welfare technology optimally (Multimedia Appendix 1, question 3). The chi square test showed no effect of gender but a strong association with profession/role (*P*<.001). Approximately half of the IT staff (13/29, 45%), chief medical nurses responsible (35/78, 45%), and OT/PTs (53/103, 51.4%) completely disagreed with the statement that their workplace optimized welfare technology.

To learn more about the respondents' attitudes and perceptions of technology (Multimedia Appendix 1, question 4), the respondents indicated their level of interest in technology. Here, some gender and age differences emerged. Overall, 191/336 (49.1%) of the women responded that "they use technology when most people do," whereas 30/53 (57%) of the men answered that they "like new technology and use technological solutions before most people."

The distribution of responses showed that most 18-24-year-old respondents (3/4, 75%) answered "I like new technology and use technological solutions before most people," whereas none of the 65-74-year-old respondents answered that they "love new technology." Considering profession/role, 16/29 (55%) of IT personnel, 33/78 (42%) of chief responsible rehabilitation specialists, and 49/103 (47.5%) of those in the "others" category responded that they "like new technology and use technological solutions before most people." In the groups of specialist dementia nurses (36/51, 71%), OT/PTs (53/103, 51.4%), and chief responsible rehabilitation specialists (15/28, 54%), the majority answered that "they use new technology when most people do."

Encouragement, Exploration, and Experimentation With Welfare Technology

The next question examined how the personnel perceived the importance to their management or closest superior of their working with welfare technology (Multimedia Appendix 1, question 5). The chi square test results showed no effect of gender and age but a strong association between profession/role and encouragement from management to use welfare technology (*P*=.01). Most respondents working in IT (26/29, 90%) reported that they were encouraged by their managers to test and use welfare technology, as did the "other" group of professionals such as eHealth strategists and unit managers (82/103, 79.6%). The results showed almost the same pattern among the other professionals, including the chief medical nurses (yes: 57/78, 73%; no: 21/78, 27%) and chief responsible rehabilitation specialists (yes: 22/28, 79%; no: 6/28, 21%).

The questionnaire then asked how much the respondents experimented with and explored welfare technology in their everyday work (Multimedia Appendix 1, question 6). The chi square test results showed an association between gender and experimentation with new welfare technology (*P*<.001). Over half of the women (186/335, 55.5%) answered that they do "not very often" experiment with welfare technology, whereas among the male respondents, only 19/53 (36%) answered that they "do not very often" experiment. The chi square test showed an independence between age and experimentation of welfare technology in daily work. There were associations between experimentation with welfare technology and the respondents' profession/role (*P*=.02). The majority of the personnel responded that they never explored or experimented with new welfare technology solutions (IT staff: 12/29, 41%; chief responsible medical nurses: 46/78, 59%; chief responsible rehabilitation specialists: 14/28, 50%; OT/PTs: 31/102, 30.3%; specialist dementia nurses: 36/51, 71%; "others": 45/103, 43.7%).

Two questions assessed whether the respondents explored and experimented with new welfare technology solutions with management and clients/patients in their everyday work (see Multimedia Appendix 1, questions 7 and 8). The results showed that more men than women experimented and explored welfare technology with management. Neither age nor work experience had any association with opportunities to experiment and explore welfare technology with management.

The chi square test also showed an association of profession/role with experimentation and exploring welfare technology with management. The IT staff reported that they occasionally (10/29, 35 %) or never (11/29, 38%) experimented and explored welfare technology with management. Among the chief responsible medical nurses, 35/78 (45%) chose the "never" option. The majority of OT/PTs and specialist dementia nurses gave almost the same answer. In response to a question concerning whether the respondents experimented and explored welfare technology solutions in collaboration with potential end users such as clients/patients, the majority (184/336, 54.8%) of the female participants responded "never," whereas 71 (21.1%) responded "yes, sometimes." Among the male participants, 23/53 (43%) answered "never." Respondents in patient-related professions, including the specialist dementia nurses and the chief responsible medical nurses, answered that they never experimented. The results also showed that professionals with longer work experience (more than 20 years) largely did not experiment and explore welfare technology with patients (91/140, 65.0%).

The next question examined whether the respondents experienced any problems in purchasing and exploring welfare technology in their profession/role (Multimedia Appendix 1, question 9). Both women (53%) and men (53%) reported experiencing such difficulties. The chi square test showed independence (*P*=.95) from gender; however, more than half the respondents indicated that they experienced problems. A clear association emerged with profession/role (*P*=.07). The majority of the IT personnel (19/29, 66%) answered that they had no such problems, whereas the majority of chief responsible medical nurses (41/78, 53%), OT/PTs (67/103, 65.0%), and specialist nurses in dementia care (26/51, 51%) did encounter such difficulties (see Table 2).

The next question asked whether the respondents regularly evaluated potential new welfare technology for their elder care organizations (Multimedia Appendix 1, question 10). These results showed independence of gender, age, and work experience. However, the majority of both the men (55%) and the women (62%) answered that they never evaluated new welfare technology. The results showed an association between profession/role and regular evaluation of new welfare technology (*P*=.02), with the majority reporting that they did not evaluate and test the suitability of new welfare technology (Table 3).

Responses to the question "Do you evaluate and follow up the welfare technology that is implemented?" were independent of gender. However, the association between profession/role and continuous evaluation of existing welfare technology was strong (*P*=.02), with the majority of participants answering that they do evaluate the welfare technology devices and solutions implemented in their organizations (Table 4).

Table 2. Distribution of responses regarding experimentation with and purchase of new welfare technology in relation to profession/role (N=393).

Profession/role	N	Yes, n (%)	No, n (%)
Information technology staff	29	10 (34.5)	19 (65.5)
Chief medical nurse responsible	78	41 (52.3)	37 (47.4)
Chief rehabilitation officer responsible	28	14 (50.0)	14 (50.0)
Occupational therapist/physiotherapist	103	67 (65.0)	36 (35.0)
Specialist dementia nurse	51	26 (51.0)	25 (49.1)
Other (manager/electronic health strategist)	103	52 (50.5)	51 (50.1)

Table 3. Distribution of responses regarding continuous evaluation of potential welfare technology for future deployment in elder care in relation to profession/role (N=393).

Profession/role	N	Yes, n (%)	No, n (%)
Information technology staff	29	13 (44.8)	16 (55.2)
Chief medical nurse responsible	78	39 (50.0)	39 (50.0)
Chief rehabilitation officer responsible	28	11 (39.1)	17 (60.7)
Occupational therapist/physiotherapist	103	27 (26.2)	76 (73.8)
Specialist dementia nurse	51	17 (33.3)	34 (66.7)
Other (manager/electronic health strategist)	103	45 (43.7)	58 (56.3)

Table 4. Distribution of responses regarding continuous evaluation of implemented welfare technology in relation to profession/role (N=393).

Profession/role	N	Yes, n (%)	No, n (%)
Information technology staff	29	23 (79.3)	6 (20.7)
Chief medical nurse responsible	78	57 (73.1)	21 (26.9)
Chief rehabilitation officer responsible	28	22 (78.6)	6 (21.4)
Occupational therapist/physiotherapist	103	57 (55.3)	46 (44.7)
Specialist dementia nurse	51	38 (74.5)	13 (25.5)
Other (manager/electronic health strategist)	103	75 (72.8)	28 (27.1)

Procurement

The respondents' involvement in the decision making related to buying and procuring welfare technology was also investigated (Multimedia Appendix 1, question 12). The chi square analysis showed that even though the majority of the participants were women, more men (30/53, 57%) were involved in the procurement process for welfare technology devices and solutions than women (98/336, 29.2%) (*P*<.001).

More than half of the chief responsible medical nurses (41/78, 53%) were involved in decision making related to buying and procurement, whereas the chief responsible rehabilitation officers (4/28, 14%), OT/PTs (7/103, 6.8%), and specialist nurses in dementia (5/51, 10%) reported little involvement. Most of the IT staff (27/29, 93%) were involved in welfare technology procurement decisions, but those who worked closely with the patients, such as the OTs (who prescribe most of the welfare technology), specialist dementia nurses, and chief rehabilitation officers responsible, had no such involvement.

Discussion

Principal Findings

The aim of this study was to explore and describe experiences of working with welfare technology in municipal elder care across gender, age, and profession. Some of the most interesting results are highlighted as follows. There was an overall positive attitude among the respondents toward digital transformation but the process was also perceived as too slow. Respondents are encouraged by management but never explore welfare technology with their managers. Moreover, elder care organizations are perceived to neither optimize nor experiment with new welfare technology. A significant decision-making and gender aspect was revealed, demonstrating issues with emancipation and welfare technology implementation. Finally, one of the key results from this study is that young personnel want to speed up the digital transformation.

Most of the respondents in this cross-sectional, explorative study were women, reflecting the nature of health care as a female-dominated occupation. In the gender distribution, the majority of IT staff and the group of “others” (eg, eHealth strategists and managers) were men. Women constituted the majority in the professions working closely with patients. The IT staff had less work experience (less than 4 years), whereas the other groups, especially the specialist dementia nurses, had far longer work experience. These demographic features of the sample are important to bear in mind to contextualize the discussion below.

Positive But Slow Digital Transformation

In general, municipal elder care personnel viewed the deployment of new technology in their work positively. Society, media, and political discourse all express the expectation that welfare technologies offer potential solutions to the demographic challenges and shortages in health care staff, and that they could have positive effects on the personnel implementing the welfare technology [4,27]. In terms of the participants’ perceptions of the speed of digital transformation, the results showed that both men and women perceive the technological developments in elder care as being “too slow.” These results confirm the high expectations for welfare technology in society and demonstrate the reality that progress is very slow in everyday elder care practices.

Personnel Are Encouraged to Explore and Experiment With Welfare Technology But Never Do

Other interesting results that emerged from the questionnaire concerned encouragement, experimentation, and exploration with welfare technology. The high level of encouragement from management that the different professional groups in municipal elder care experienced would seem to reflect the Swedish government’s positive views on the use of welfare technology [4]. This study shows that most management encouraged personnel to experiment with and explore welfare technology. However, most of the participants had never explored welfare technology with their managers. Based on these findings, it can be concluded that personnel (eg, eHealth strategist) make decisions on new technology purchases without management’s involvement in experimentation with the new welfare technology. Alternatively, it can be assumed that some welfare technology is procured and purchased without being tested. Further studies are needed to explore these aspects.

Elder Care Personnel Neither Optimize Nor Experiment With New Welfare Technology

Participants gave negative responses as to whether they perceived their workplace as optimizing the use of welfare technology. Those working closely with patients, including OT/PTs and specialist dementia nurses, perceived that the elder care organizations could optimize the use of welfare technology to a greater extent. However, some of the participants also responded that technology had changed health care organizations in negative ways. These findings might indicate a negative reaction to the rapid technological change in society or the overburdening of these professions/roles with responsibilities in their everyday work.

The association of profession/role and experimentation with new welfare technology was strong, with a majority of the respondents answering that they never experimented. Interestingly, the participants felt encouraged to experiment with and explore welfare technology but never did so with either management or patients. These findings could indicate that patients in need of elder care have little influence on the kinds of welfare technology recommended and implemented. Substantial research has shown the importance of patient participation in the procurement process for assistive and welfare technologies [28,29-33]. Evidence supports the conclusion that patients use assistive technology and welfare technology more if they are involved in the entire process of its procurement [34].

Decision Making and Gender

More than half the female participants reported that they “use technology when most people do,” whereas the majority of men responded that they “like new technology and use technology before most people.” These findings support earlier research showing that men have greater confidence in using new technologies than women [35]. Studies on gender and technology have shown that the introduction of new technologies tends to degrade women’s work and separate it from men’s work [36], while women tend to be replaced by men during technological change in the workplace [37]. As welfare technology enters the generally female domain of elder care, men seem to carry the professional role (IT strategist) of introducing the welfare technology into municipal elder care. The present results showed that men have a more positive attitude toward and curiosity regarding new technologies than women. This may result from their role in the organization and may not be due to a gender effect itself. However, this finding raises the question as to whether the introduction of technology into municipal elder care will lead to an increase in male care professionals.

Considering welfare technology procurement, the results showed that although most of the participants were women, more male respondents were involved in the decision making than women.

These results confirm that technology is often constructed as a male domain [38,39], and that it is not neutral but is rather shaped by social values and norms [40,41]. These results indicate that the male respondents had more power than the female respondents in the decision-making process regarding the deployment and use of technological innovations in the elder care sector.

Historically, women have experienced the negative effects of hierarchical structures of power and technology deployment [37,38,41-43]. The digital transformation has, and will continue to have, major impacts on many aspects of everyday and professional life, including gender equality. Digital transformation will affect political participation, the nature of the labor market, and interactions with friends and colleagues, opening opportunities for dealing with gender differences that affect both women and men, albeit often differently [43]. Digital transformation has the potential to either reproduce and maintain gender expectations and power structures or to advance a more equal society. Digitization and technological innovations take place not in a vacuum but in interactions with social, cultural, economic, and political factors [44].

Personnel in municipal elder care are shaped by their working conditions, and, unfortunately, our results indicate that welfare technology maintains and reproduces gender expectations and power structures in municipal elder care. For example, the male participants had much higher levels of power in the decision-making process by which new technological solutions were selected for implementation in municipal elder care.

Emancipation and Welfare Technology Implementation

Almost all of the IT staff and the chief rehabilitation officers responsible were involved in procurement decisions, but among those working closely with patients, OT/PTs and specialist dementia nurses did not participate in these decisions. These results suggest that those with strong power in the procurement process do not work closely with patients or even regularly meet them, while those who work directly with patients daily and implement the new technological solutions have little power and do not control what solutions or devices are procured. The consequence might be a resistance to technology among frontline care personnel and patients in municipal elder care [10,15,45,46]. Recent studies have analyzed aspects of participation among different care professionals when welfare technologies have been implemented [12,17,19,20,45-49]. These studies have shown that the work of the individual care personnel is crucial in technology deployment in elder care and that technology innovation also changes the dynamics within practices [21]. However, this research has generally failed to acknowledge the participation or lack of participation by care professionals in the initial exploration and experimentation with welfare technology before procurement and implementation, which requires further study. Research on well-being in the workplace has shown that having control and a voice in the workplace are key factors for well-being and satisfaction at work [50-53]. A lack of a voice on which welfare technology solutions are implemented may therefore negatively affect those working closely with patients, consequently increasing their work-related stress and administrative load [54].

Young Personnel Want to Speed Up the Digital Transformation

Our findings indicate that not only gender but also age affect the perception of welfare technology. The association between age and perceptions of the speed of digital transformation shows the importance of age as a variable. All of the 18 to 24-year-old participants thought that the speed was "too slow," whereas more than half of the 65 to 74-year-old participants thought that it was moving at the "right pace." These findings indicate that the perception that digital transformation is proceeding at the right pace increases with age. Prensky [52] defines "digital natives" as those who have used and been surrounded by technology their whole lives and "digital immigrants" as those who learn and adopt to technology and have had experience of doing similar tasks without technology. Prensky [52] proposed that digital natives have significant advantages when using technology. This was confirmed by our study results in which the younger respondents declared that they "like new technology and use it before most people do."

Limitations

One limitation of this study was that the questionnaire was sent to municipal registrars who passed it on to potential participants working with welfare technology, resulting in the participants occupying different levels and roles in organizations, and included registered nurses, OT/PTs, managers, and IT strategists. It would have been interesting to determine the level at which the individual participants worked in terms of how much decision-making power they had for buying and implementing welfare technology; however, this was not the aim of the study. Another limitation was that the questionnaire was anonymous, and therefore the geographic spread of the participants in Sweden was unclear. However, a total of 393 responses were received and Sweden has 290 municipalities, suggesting a good geographic spread.

Conclusion

The findings show that various professions working in municipal elder care hold very positive views on new technologies. However, there are both gender and age differences in attitudes toward technology (for the use of technology in general and in the workplace) and in participation in decision making for the procurement of new welfare technologies. The consequences of this might include personnel's resistance to welfare technology in municipal health care [46], which would affect clients and their potential participation in welfare technology. Therefore, the people working closely with patients who are expected to implement the new technologies need to be more involved in the procurement process.

Among other findings, a majority of the IT staff were men who had worked in municipal elder care for less than 4 years, and most of these participants were involved in welfare technology procurement decisions. These findings imply that men with little work experience or knowledge of elder care carry high levels of power and responsibility for decisions that will be executed by the female-dominated groups of elder care personnel. The decisions of these few could have a major influence on future elder care as digital transformation changes how work is carried out in municipal elder care.

Acknowledgments

The research presented in this paper was funded by National Network of Assistive Technology Managers (Nationella Hjälpmedelschefsnätverket) and the School of Health, Care and Social Welfare, Mälardalen University, Sweden. We thank Professor Britt Östlund who secured the funding for this research. We are also truly grateful to Sylvia Olsson, Emmie Wahlström, and Carina Loeb for their perceptive statistical input and discussions, and to all of the respondents who volunteered to participate in the study. We also want to thank Natasha Layton for her skillful input.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire on ongoing digital transformation in municipal elder care. WT: welfare technology.

[DOCX File , 22 KB - jmir_v22i6e15450_app1.docx]

Multimedia Appendix 2

Associations among variables for analysis. WT: welfare technology.

[DOCX File , 15 KB - jmir_v22i6e15450_app2.docx]

References

1. Hofmann B. Ethical challenges with welfare technology: a review of the literature. Sci Eng Ethics 2013 Jun 5;19(2):389-406. [doi: 10.1007/s11948-011-9348-1] [Medline: 22218998]
2. Cook A, Hussey S. Assistive technologies: principles and practice. St Louis: Mosby; 2002.
3. Scherer MJ. Technology adoption, acceptance, satisfaction and benefit: integrating various assistive technology outcomes. Disabil Rehabil Assist Technol 2017 Jan 04;12(1):1-2. [doi: 10.1080/17483107.2016.1253939] [Medline: 27915579]
4. Wickstrom G, Regner A, Micko L. Govérnent Offices of Sweden. Stockholm: Ministry of Health and Social Affairs; 2025. Vision for eHealth 2025 URL: https://www.government.se/information-material/2016/08/vision-for-ehealth-2025/ [accessed 2020-05-25]
5. National Board of Health and Welfare (Socialstyrelsen). 2019 Jan 01. E-health and Welfare Technology within Swedish Municipalities URL: https://www.government.se/49b757/contentassets/65d092eda66b437e9cb5e54bb4e41319/national-ehealth---the-strategy-for-accessible-and-secure-information-in-health-and-social-care-s2011.023 [accessed 2020-05-25]
6. Bergschöld JM. Domesticating Homecare Services; Vehicle Route Problem Solver Displaced. Nordic J Sci Technol Stud 2016 Dec 29;4(2):41. [doi: 10.5324/njsts.v4i2.2184]
7. Frennert S. Lost in digitalization? Municipality employment of welfare technologies. Disabil Rehabil Assist Technol 2019 Aug 28;14(6):635-642. [doi: 10.1080/17483107.2018.1496362] [Medline: 30264649]
8. Frennert S, Östlund B. Narrative Review: Technologies in Eldercare. Nordic J Sci Technol Stud 2018 Sep 14;6(1):21-34. [doi: 10.5324/njsts.v6i1.2518]
9. Stranz A, Szebehely M. Organizational trends impacting on everyday realities: the case of Swedish eldercare. In: Christensen K, Pilling D, editors. The Routledge Handbook of Social Care Work Around the World. New York: Routledge; 2017:46-58.
10. Nilsen ER, Dugstad J, Eide H, Gullslett MK, Eide T. Exploring resistance to implementation of welfare technology in municipal healthcare services - a longitudinal case study. BMC Health Serv Res 2016 Nov 15;16(1):657-657 [FREE Full text] [doi: 10.1186/s12913-016-1913-5] [Medline: 27846834]
11. Ertner M. Different generalizations of the elderly in design of welfare technology. STS Encounters 2016;8(1):1-28 [FREE Full text]
12. Hansen AM, Grosen SL. Transforming Bodywork in Eldercare with Wash-and-dry Toilets. Nordic J Work Life Stud 2019 Mar 02;9(S5):49-67. [doi: 10.18291/njwls.v9is5.112689]
13. Sjoberg P, Olsson S, Larsen C. RISE. 2019 Mar 18. Lagesrapport om valfardsteknik till Socialstyrelsen URL: https://www.sics.se/sites/default/files/pub/sics_till_socialstyrelsen_lagesrapport_valfardsteknik.pdf [accessed 2019-06-04]
14. Felding S, Schwennesen N. Digital meeting places: creating intimate and safe connections with digital means. 2019 Jun 11 Presented at: The digitally engaged patient; June 11-12, 2019; Copenhagen, Denmark p. 16-17.
15. Sävenstedt S, Sandman P, Zingmark K. The duality in using information and communication technology in elder care. J Adv Nurs 2006 Oct;56(1):17-25. [doi: 10.1111/j.1365-2648.2006.03975.x] [Medline: 16972914]

16. Kamp A, Obstfelder A, Andersson K. Welfare Technologies in Care Work. Nordic J Work Life Stud 2019 Mar 02;9(S5):1-12. [doi: 10.18291/njwls.v9is5.112692]
17. Ertner M. Enchanting, Evoking, and Affecting: The Invisible Work of Technology Implementation in Homecare. Nordic J Work Life Stud 2019 Mar 02;9(S5):33-47. [doi: 10.18291/njwls.v9is5.112690]
18. Cozza M. Interoperability and convergence for welfare technology. 2018 Jul 15 Presented at: International Conference on Human Aspects of IT for the Aged Population; July 15-20, 2018; Las Vegas, NV. [doi: 10.1007/978-3-319-92037-5_2]
19. Zander V, Johansson-Pajala R, Gustafsson C. Methods to evaluate perspectives of safety, independence, activity, and participation in older persons using welfare technology. A systematic review. Disabil Rehabil Assist Technol 2020 May;15(4):373-393. [doi: 10.1080/17483107.2019.1574919] [Medline: 30786779]
20. La Cour A, Højlund H. Untimely Welfare Technologies. Nordic J Work Life Stud 2019 Mar 02;9(S5):69-87. [doi: 10.18291/njwls.v9is5.112688]
21. Stokke R. Older People Negotiating Independence and Safety in Everyday Life Using Technology: Qualitative Study. J Med Internet Res 2018 Oct 19;20(10):e10054 [FREE Full text] [doi: 10.2196/10054] [Medline: 30341049]
22. Polit D, Beck C. Nursing Research: Principles and Methods. Philadelphia, USA: Lippincott Williams & Wilkins; 2004.
23. Frennert S, Baudin K. The concept of welfare technology in Swedish municipal eldercare. Disabil Rehabil 2019 Sep 10:1-8. [doi: 10.1080/09638288.2019.1661035] [Medline: 31503509]
24. Streiner DL, Norman GR, Cairney J. Health measurement scales: A practical guide to their development and use. Oxford, UK: Oxford University Press; Apr 1992:294.
25. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006 Feb;119(2):166.e7-16. [doi: 10.1016/j.amjmed.2005.10.036] [Medline: 16443422]
26. Swedish Council of Research. Stockholm, Sweden: Swedish Council of Research; 2017. Good Research Practice URL: https://www.vr.se/download/18.5639980c162791bbfe697882/1555334908942/Good-Research-Practice_VR_2017.pdf [accessed 2020-05-25]
27. Søndergård D, Hadnagy J, Danielsson A, Rolfer B, Rottem KL. Välfärdsteknik handlar inte om teknik utan om människor-tekniksprång i nordisk demensvård. In: NVC Nordens Välfärdcenter/Nordic Centre for Welfare and Social Issues. Stockholm, Sweden: Nordens Välfärdcenter/Nordic Centre for Welfare and Social Issues; 2017:1-19.
28. MacLachlan M, Banes D, Bell D, Borg J, Donnelly B, Fembek M, et al. Assistive technology policy: a position paper from the first global research, innovation, and education on assistive technology (GREAT) summit. Disabil Rehabil Assist Technol 2018 Jul 23;13(5):454-466. [doi: 10.1080/17483107.2018.1468496] [Medline: 29790393]
29. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
30. de Witte L, Steel E, Gupta S, Ramos VD, Roentgen U. Assistive technology provision: towards an international framework for assuring availability and accessibility of affordable high-quality assistive technology. Disabil Rehabil Assist Technol 2018 Jul 09;13(5):467-472. [doi: 10.1080/17483107.2018.1470264] [Medline: 29741965]
31. Andrich R, Mathiassen N, Hoogerwerf E, Gelderblom GJ. Service delivery systems for assistive technology in Europe: An AAATE/EASTIN position paper. Technol Disab 2013 Aug 06;25(3):127-146. [doi: 10.3233/tad-130381]
32. MacLachlan M, Scherer MJ. Systems thinking for assistive technology: a commentary on the GREAT summit. Disabil Rehabil Assist Technol 2018 Jul 17;13(5):492-496. [doi: 10.1080/17483107.2018.1472306] [Medline: 29772950]
33. Lersilp S, Putthinoi S, Okahashi S. Information Management for the Assistive Technology Provision in Community: Perspectives of Local Policymakers and Health Service Providers. Occup Ther Int 2018;2018:809283. [doi: 10.1155/2018/8019283] [Medline: 29681786]
34. Löfqvist C, Slaug B, Ekström H, Kylberg M, Haak M. Use, non-use and perceived unmet needs of assistive technology among Swedish people in the third age. Disabil Rehabil Assist Technol 2016 Sep 19;11(3):195-201. [doi: 10.3109/17483107.2014.961180] [Medline: 25238550]
35. Broos A, Roe K. The digital divide in the playstation generation: Self-efficacy, locus of control and ICT adoption among adolescents. Poetics 2006 Aug;34(4-5):306-317. [doi: 10.1016/j.poetic.2006.05.002]
36. Glenn EN, Feldberg RL. Degraded and Deskilled: The Proletarianization of Clerical Work. Social Problems 1977 Oct;25(1):52-64. [doi: 10.1525/sp.1977.25.1.03a00060]
37. Sommestad L. In: Osterud G, editor. Agrarian Women, the Gender of Dairy Work, and the Two-breadwinner Model in the Swedish Welfare State. New York: Routledge; 2019.
38. Oldenziel R. Boys and Their Toys: The Fisher Body Craftsman's Guild, 1930-1968, and the Making of a Male Technical Domain. Technol Cult 1997 Jan;38(1):60. [doi: 10.2307/3106784]
39. Durack K. Gender, Technology, and the History of Technical Communication. Tech Commun Quart 1997 Jul;6(3):249-260. [doi: 10.1207/s15427625tcq0603_2]
40. Akrich M. The description of technical objects. In: Bikjer WE, Law J, editors. Shaping technology/building society, studies in sociotechnical change. Cambridge, MA: MIT Press; 1992:205-224.
41. Faulkner W. The technology question in feminism. Womens Stud Int Forum 2001 Jan;24(1):79-95. [doi: 10.1016/s0277-5395(00)00166-7]

42. Gemzöe L. Feminism. Stockholm, Sweden: Bilda; 2014.
43. Cockburn C, Ormrod S. Gender and Technology in the Making. London: SAGE; 1993:8147.
44. Berger P, Luckmann T. The social construction of reality. UK: Penguin Books Ltd; Mar 28, 1991:256.
45. Safi S, Thiessen T, Schmailzl KJ. Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study. JMIR Res Protoc 2018 Dec 04;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
46. Chen SC. Technological Health Intervention in Population Aging to Assist People to Work Smarter not Harder: Qualitative Study. J Med Internet Res 2018 Jan 04;20(1):e3 [FREE Full text] [doi: 10.2196/jmir.8977] [Medline: 29301736]
47. Kamp A, Hansen AM. Negotiating Professional Knowledge and Responsibility in Cross-sectoral Telemedicine. Nordic J Work Life Stud 2019 Mar 02;9(S5):13-32. [doi: 10.18291/njwls.v9is5.112691]
48. López Gómez D. Little arrangements that matter. Rethinking autonomy-enabling innovations for later life. Technol Forecast Social Change 2015 Apr;93:91-101. [doi: 10.1016/j.techfore.2014.02.015]
49. Sánchez-Criado T, López D, Roberts C, Domènech M. Installing Telecare, Installing Users. Sci Technol Hum Values 2014 Jan 09;39(5):694-719. [doi: 10.1177/0162243913517011]
50. Landsbergis PA. Occupational stress among health care workers: A test of the job demands-control model. J Organiz Behav 1988 Jul;9(3):217-239. [doi: 10.1002/job.4030090303]
51. Brandtzæg P, Følstad A, Heim J. Enjoyment: lessons from Karasek. In: Blythe M, Monk A, editors. Funology 2: From Usability to Enjoyment. Switzerland: Springer Nature; 2018:331-341.
52. Prensky M. Digital natives, digital immigrants. From On the Horizon. MCB Univ Press 2011;9(5):1-6.
53. Taris T, Leisink P, Schaufeli W. Applying occupational health theories to educator stress. In: McIntyre T, McIntyre S, Francis D, editors. Educator Stress: Aligning Perspectives on Health, Safety and Well-Being. Switzerland: Springer Cham; 2017:280.
54. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]

Abbreviations

eHealth: electronic health
IT: information technology
OT: occupational therapist
PT: physiotherapist

Edited by G Eysenbach; submitted 13.07.19; peer-reviewed by AS Aslam, W Zhang; comments to author 22.10.19; revised version received 18.02.20; accepted 29.03.20; published 16.06.20.

Please cite as:
Baudin K, Gustafsson C, Frennert S
Views of Swedish Elder Care Personnel on Ongoing Digital Transformation: Cross-Sectional Study
J Med Internet Res 2020;22(6):e15450
URL: http://www.jmir.org/2020/6/e15450/
doi:10.2196/15450
PMID:32543444

Original Paper

The Demographic Representativeness and Health Outcomes of Digital Health Station Users: Longitudinal Study

Leah Flitcroft[1], BHSc, MSc; Won Sun Chen[1], BSc, MSc, DPhil; Denny Meyer[1], BScHons, MBL, DBL

[1] Faculty of Health, Arts & Design, Swinburne University of Technology, Hawthorn, Australia

Corresponding Author:
Won Sun Chen, BSc, MSc, DPhil
Faculty of Health, Arts & Design
Swinburne University of Technology
Level 9
PO Box 218
Hawthorn, 3122
Australia
Phone: 61 3 9214 8437
Email: wchen@swin.edu.au

Abstract

Background: Digital health stations offer an affordable and accessible platform for people to monitor their health; however, there is limited information regarding the demographic profile of users and the health benefits of this technology.

Objective: This study aimed to assess the demographic representativeness of health station users, identify the factors associated with repeat utilization of stations, and determine if the health status of repeat users changed between baseline and final health check.

Methods: Data from 180,442 health station users in Australia, including 8441 repeat users, were compared with 2014-2015 Australian National Health Survey (NHS) participants on key demographic and health characteristics. Binary logistic regression analyses were used to compare demographic and health characteristics of repeat and one-time users. Baseline and final health checks of repeat users were compared using McNemar tests and Wilcoxon signed rank tests. The relationship between the number of checks and final health scores was investigated using generalized linear models.

Results: The demographic profile of SiSU health station users differs from that of the general population. A larger proportion of SiSU users were female (100,814/180,442, 55.87% vs 7807/15,393, 50.72%), younger (86,387/180,442, 47.88% vs 5309/15,393, 34.49% aged less than 35 years), and socioeconomically advantaged (64,388/180,442, 35.68% vs 3117/15,393, 20.25%). Compared with NHS participants, a smaller proportion of SiSU health station users were overweight or obese, were smokers, had high blood pressure (BP), or had diabetes. When data were weighted for demographic differences, only rates of high BP were found to be lower for SiSU users compared with the NHS participants (odds ratio [OR] 1.26; *P*<.001). Repeat users were more likely to be female (OR 1.37; *P*<.001), younger (OR 0.99; *P*<.001), and from high socioeconomic status areas—those residing in socioeconomic index for areas quintiles 4 and 5 were significantly more likely to be repeat users compared with those residing in quintile 1 (OR 1.243; *P*<.001 and OR 1.151; *P*<.001, respectively). Repeat users were more likely to have a higher BMI (OR 1.02; *P*<.001), high BP (OR 1.15; *P*<.001), and less likely to be smokers (OR 0.77; *P*<.001). Significant improvements in health status were observed for repeat users. Mean BMI decreased by 0.97 kg/m2 from baseline to final check (z=−14.24; *P*<.001), whereas the proportion of people with high BP decreased from 15.77% (1080/6848) to 12.90% (885/6860; χ^2_1=38.2; *P*<.001). The proportion of smokers decreased from 11.91% (1005/8438) to 10.13% (853/8421; χ^2_1=48.4; *P*<.001). Number of repeat health checks was significantly associated with smoking status (OR 0.96; *P*<.048) but not with higher BP (*P*=.14) or BMI (*P*=.23).

Conclusions: These findings provide valuable insight into the benefits of health stations for self-monitoring and partially support previous research regarding the effect of demographics and health status on self-management of health.

(J Med Internet Res 2020;22(6):e14977) doi:10.2196/14977

KEYWORDS

population health; health behavior; health technology; eHealth; health status

Introduction

Background

Technological advances in recent years have changed the way consumers access health care and enabled the use of a range of digital mechanisms for self-monitoring of health, including mobile phone apps, wearable trackers, and web-based monitoring systems [1,2]. Although self-monitoring in general has been shown to be effective in the management of health risk factors and chronic disease and in increasing self-efficacy in disease management [3-5], limited research has been conducted to date regarding the long-term benefits of using digital technologies to self-monitor health [6,7]. Some benefits have been identified in terms of supporting behavior change [7,8], promoting weight loss [7,9,10], increasing physical activity [7,8,10,11], assisting with smoking cessation [7,10], and improving self-management of chronic disease [7,12], particularly in those with more serious chronic illness [6]; however, findings to date have been inconsistent.

The increased utility of digital self-monitoring technology over traditional paper-based tracking, in terms of it enabling the easy collection and exchange of health-related information between consumers and health care providers, has also been supported by a number of authors [13-15]. There is some evidence that data collected through such technologies is more reliable than data collected via other means such as manual measurement, self-report, and in some cases medical professionals [16-18]. The mechanisms behind these findings are varied and include the competence of the general public in using and interpreting medical instrumentation such as sphygmomanometers [19], social desirability associated with self-reports of disease and risk factors [16,20], and a phenomenon known as *white coat syndrome*, in which blood pressure (BP) and heart rate are artificially elevated in a clinical setting [21,22].

The uptake and utilization of health monitoring devices and apps appear to be influenced by a range of social, health-related, and demographic factors, including age [23-25], health status [25-27], and socioeconomic status (SES) [23,25,28]. However, it is possible that the influences of such characteristics may be dependent on the type of technology in question [23].

Given the potential benefits to both consumers and health professionals, and the rapid increase in the use of technology in the health arena, there is some concern surrounding the influence of sociodemographic characteristics on access to digital technologies [23,29]. Such influences may result in the potential for certain disadvantaged subgroups to be *left behind* in the digital health age and unable to benefit from the potential of such technologies [23]. Older age and socioeconomic disadvantage, in particular, are often associated with barriers to the utilization of digital health apps, due to poor levels of electronic health literacy, prohibitive costs, and limited access to the internet [23,30].

The SiSU Health Group is a health and wellness company that aims to ease the effects of lifestyle-related diseases on global health care systems through the use of technology. Their health check stations, installed in a number of locations across Australia, offer an affordable and accessible platform to help people live a healthier life. SiSU health stations are free of charge for all Australians aged 16 years and older, providing an alternative method of enabling the general population to monitor their health and access relevant health information. These stations have the potential to reach consumers who face barriers to the utilization of digital health technologies [29-31], particularly those facing economic barriers to such technologies. Although several studies have been published that investigate the determinants of health kiosk utilization, findings are varied. Socioeconomic factors including income, employment status, country of birth, gender, and age have all been found to influence the utilization of health kiosks to varying degrees [29-33]; however, the results are inconsistent.

There is some evidence to demonstrate that the utilization of health kiosks to access health information can lead to increased screening rates, improved health literacy, and a reduction in the burden on medical services [33]. The majority of these studies, however, investigated the utilization of kiosks that provide health information only, with very few including kiosks that enable health measurement and self-monitoring [29-32] such as the SiSU health station. Limited information is available regarding the potential benefits of health kiosk utilization for self-monitoring in terms of improved health status and a reduction in health risk behaviors. For this reason, this study aimed to provide insight into the demographic and health-related characteristics that are associated with the utilization of health kiosks for self-monitoring and identify any observed improvements in the health status of users over time, using data obtained from SiSU health stations. These stations are installed in a number of pharmacies, retail outlets, and workplaces throughout Australia and the United Kingdom. In Australia, the majority are currently located in Priceline Pharmacies.

Objectives

The SiSU wellness health check station collects data on a range of self-reported and machine-measured health indicators, including diabetes status, physical activity levels, waist circumference, dietary practices, heart rate, BP, weight, BMI, and body fat percentage. The SiSU station is designed to be a vertical space with mobile 3G connection or access to a private Wi-Fi network. Users engage with the station for approximately 7 min to answer a series of questions and to provide various health-related measures, before they receive immediate feedback about their health status on the screen of the station. Users of the health stations are able to monitor their progress and health changes over time by connecting their health check station profile to a free app developed by SiSU Health Group and downloadable from Google Play or the iTunes store. The stations and associated apps are intended to assist consumers by allowing them to monitor their health status over time and providing alerts to consumers when follow-up with a general practitioner is recommended [34].

There is a lack of evidence regarding the health benefits of kiosks that allow self-monitoring of health outcomes and inconsistent evidence regarding the demographic of users. This study, therefore, aimed to determine if the users of SiSU health stations in Australia differ from the general population in terms of demographics and health status by comparing SiSU health station users with participants of the 2015 National Health Survey (NHS); investigate the demographic and health-related characteristics that are associated with repeated utilization of the SiSU health stations in Australia; and identify if the health status of repeat users of SiSU health stations in Australia improved from baseline (at their first check) to their final health check.

Methods

Study Sample

This study uses data collected from 192 SiSU health check stations installed across Australia, for the period October 28, 2017, to June 27, 2018. Due to resource limitations, data for health check stations in the United Kingdom were not considered in this study. This resulted in a total of 271,151 records pertaining to males and females aged 16 years and above in Australia. Users reporting a pregnancy at one or more of their health station checks were removed from the dataset (n=3315). The majority (266,813/271,151, 98.40%) of data were obtained from health check stations installed in Priceline Pharmacies across Australia.

For research question 1—analysis of the demographics and health status of SiSU users—and 2—investigate the demographic and health-related characteristics that are associated with repeated utilization of the SiSU health stations in Australia—only the first health checks of users were included to avoid bias in measurements introduced through potential improvements in health status as a result of self-monitoring. Therefore, records that were identified as repeat checks were excluded from the analysis, as were any users who had undertaken their first health check in the time before the study period. Invalid measurements were also identified and removed, resulting in a sample size of 180,442 records.

The demographics and health status of the users were compared with those of the participants in the 2014-2015 Australian NHS, a nationally representative survey of 19,000 people in approximately 15,000 households. NHS data are weighted to reflect sampling fractions for each respondent, ensuring that the results are representative of the general population [35]. Only NHS participants aged 16 years and above were included in this analysis, resulting in a total of 15,393 records.

For research question 3—analysis of changes in health status from baseline—only users who had undertaken 2 or more health checks were included. Users were classified as repeat users if their unique user ID appeared more than once in the dataset. Suspected shared accounts were also identified by comparing the age and gender recorded at each health check. User IDs with inconsistent entries for these variables were considered to be shared accounts and were removed from the analysis, resulting in a total of 27,522 health checks pertaining to 8441 users.

Measures

Data on the following variables were used in this study: gender, age, SES, state, BP, BMI, diabetes status, smoking status, and repeat user status.

The health station questions regarding fruit and vegetable consumption and physical activity are only asked of nondiabetic health station users, resulting in a large amount of missing data for each of these variables (64,959/180,442, 36.00% missing). Therefore, these variables were excluded from this analysis.

The format of the age variable in the NHS data, which was grouped into 5 year categories, necessitated the grouping of the SiSU health station age variable into categories for the purposes of research question 1. To reduce the number of categories, 10 year age groups were selected. For the remaining research questions, age was treated as a continuous variable.

SES was defined using the 2016 version of the index of relative advantage and disadvantage (IRSAD) under the Australian Bureau of Statistics’(ABS) socioeconomic indexes for areas (SEIFA), which summarizes a range of variables that are considered to represent relative socioeconomic advantage and disadvantage. The IRSAD ranks geographical areas on a continuum from most disadvantaged to least disadvantaged based on this summary. Areas are ranked into quintiles, where quintile 1 contains the lowest 20% of areas (most disadvantaged), quintile 2 contains the next lowest 20% of areas, and so forth, resulting in 5 equal-sized groups. SEIFA scores were allocated to SiSU health station users by matching their residential postcode to the SEIFA index. State was recorded as the state in which the person resided at the time of their health check.

BP was grouped into categories based on measured systolic and diastolic readings: a reading of ≥140/90 mm Hg was categorized as high. Similarly, BMI was calculated using a person’s measured height (m) and weight (kg), and values of ≥30.0 kg/m^2 were categorized as overweight or obese. Binary categories were chosen for these variables to enable a direct comparison of the proportion of participants with or without the respective health conditions (prevalence rates).

Diabetes status and smoking status were recorded as *yes* if a person had an affirmative response to the questions: “Do you have a current diagnosis of type 1 or type 2 diabetes?” and “Are you a current smoker?,” respectively.

Missing Data

Analysis of missing values was performed on the SiSU wellness data to determine the volume of missing data and identify factors associated with missingness. The total proportion of missing values in the dataset was 5.40% (9744/180,442), with BMI and BP having the highest proportion of missing values at 19.00% (34,284/180,442) and 18.79% (33,923/180,442), respectively. Due to the high rate of missing data, inverse probability weighting (IPW) was used to weight records according to their probability of being a complete record, according to the methodology detailed in the study by Seaman and White [36]. A binary logistic regression was conducted to calculate the IPW, with the variables gender, age, state of residence, and SES included in the model. Separate weights were calculated for each missing variable. All statistical analyses were conducted with these weights applied, using only records with complete data for the variables in each model.

Missing data for physical measurements was higher in the NHS compared with the SiSU wellness data. In the 2014-2015 NHS, physical measurements were taken for height, weight, and BP. A total of 24.29% (3740/15,393) of respondents did not have their BP measured, whereas 26.80% (4125/15,393) did not have their height, weight, or both measured. The NHS utilized imputation to estimate physical measurements for these participants.

Statistical Analysis

To assess if the users of SiSU health stations are representative of the general population, selected demographic and health characteristics of the sample were compared with the those of the Australian population using data from the 2014-2015 NHS.

Characteristics that were directly comparable between the SiSU health station users and the NHS dataset were age group, sex, SES, state of residence, BP, BMI, diabetes status, and smoking status. Demographic characteristics of the 2 groups were compared, and SiSU health station data were then weighted to account for demographic differences between the SiSU health station users and NHS participants. Using these weighted data, logistic regression was used to obtain odds ratios (ORs) to determine the magnitude of any differences between the health-related measures for SiSU health station users compared with NHS participants. In total, 4 models were created, 1 for each of the health characteristic variables—BMI, BP, diabetes status, and smoking status—with the health characteristic as the dependent variable and *group* (SiSU user or NHS participant) as the independent variable. The demographic variables age, gender, SES, and state were included as controls in these models.

Comparisons were also made between repeat and nonrepeat users of the SiSU health stations. Users were classified as repeat if they had undertaken 2 or more health checks within the time period, whereas users were considered nonrepeat if they had only undertaken 1 health check during the period or had no recorded user ID number. Binary logistic regression was performed to identify the demographic and health-related factors that predict the probability of being a repeat user. With repeat status as the dependent variable, demographic and health predictors included in the model were age, gender, SES, state, BMI, BP category, diabetes status, and smoking status. Binary logistic regression was chosen because of the binary nature of the dependent variable and the ability to introduce covariates for analysis and quantify the relationship between the dependent and independent variables in terms of ORs [37,38].

To establish if the health status of repeat users changed between their first and final health checks, the baseline (first health check) and final (last check identified in the period under study) health measurements of repeat users were compared. McNemar tests were performed for the categorical variables BP category and smoking status to identify any change in proportions between baseline and final checks, and a Wilcoxon signed rank test was performed for baseline and final BMI measurements to identify any change in mean BMI scores from baseline to final check. Both methods were chosen because of their ability to allow the comparison of related or paired samples, and both the methods are often used in research comparing pre and posttreatment measurements [39,40].

Finally, to determine if the number of health checks completed by a user affects health outcomes at their final check, binary logistic regression models were constructed for each of the health outcomes BP category and smoking status, again because of their ability to include covariates and produce ORs for the quantification of relationships. A generalized linear model was used for BMI, assuming a gamma distribution due to the skewness of the data. This is a method suggested by some authors to overcome the issue of right-skewed data while avoiding issues associated with log-retransformation [41]. Baseline measurements were included in these models to control for differences in baseline health scores, and the demographic variables gender, age, and SES were included as covariates.

A *P* value of less than .05 (2-tailed) was deemed to be statistically significant. All analyses were performed using International Business Machines SPSS version 21.

Results

Comparison of Demographic and Health Characteristics of SiSU Health Station Users and National Health Survey Participants

The demographic characteristics of SiSU health station users compared with NHS users are presented in Table 1, revealing large differences in proportions across all demographic variables under consideration. SiSU health stations users were found to be younger (86,387/180,442, 47.87% vs 5309/15,393, 34.49% aged less than 35 years; *P*<.001) and living in higher SES areas (64,388/180,442, 35.68% vs 3117/15,393, 20.25% in quintile 5; *P*<.001) compared with NHS participants. A higher proportion of SiSU health station users were female (100,814/180,442, 55.87% vs 7807/15,393, 50.72%; *P*<.001), and the proportion of SiSU users living in each state varied from that of NHS participants, with a larger proportion of SiSU health station users residing in New South Wales compared with the NHS participants (92,636/180,442, 51.34% vs 4982/15,393, 32.37%; *P*<.001).

Although detailed data are not available for Priceline Pharmacy customers, according to Priceline Pharmacies, 97.00% of their customer base is female [42]. This gender distribution varies greatly compared with that of NHS participants and SiSU health station users.

Table 1. Comparison of the demographic characteristics of SiSU health station users and National Health Survey participants.

Variable	SiSU health station users (n=180,442), n (%)	NHS[a] participants (n=15,393), n (%)	*P* value
Age group (years)			<.001
>24	42,027 (23.29)	2499 (16.23)	
25-34	44,360 (24.58)	2810 (18.26)	
35-44	26,905 (14.91)	2623 (17.04)	
45-54	23,529 (13.04)	2530 (16.44)	
55-64	22,878 (12.68)	2211 (14.36)	
65-74	15,101 (8.37)	1604 (10.42)	
≥75	6452 (3.58)	1117 (7.26)	
Gender			<.001
Male	79,628 (44.13)	7586 (49.28)	
Female	100,814 (55.87)	7807 (50.72)	
SEIFA[b] quintile			<.001
1	26,439 (14.65)	2961 (19.24)	
2	27,899 (15.46)	3019 (19.61)	
3	25,197 (13.96)	3123 (20.29)	
4	36,519 (20.24)	3172 (20.61)	
5	64,388 (35.68)	3117 (20.25)	
State			<.001
New South Wales	92,636 (51.34)	4982 (32.37)	
Victoria	23,377 (12.96)	3914 (25.43)	
Queensland	35,513 (19.68)	3039 (19.74)	
South Australia	9760 (5.41)	1113 (7.23)	
Western Australia	12,824 (7.11)	1637 (10.63)	
Tasmania	3240 (1.80)	338 (2.20)	
Northern Territory	154 (0.10)	115 (0.74)	
Australian Capital Territory	2938 (1.63)	254 (1.65)	

[a] NHS: National Health Survey.

[b] SEIFA: socioeconomic indexes for areas.

Table 2 provides a comparison of the unweighted and weighted health characteristics of SiSU health station users and NHS participants. Before weighting, SiSU health station users were generally healthier than NHS participants on the variables measured. There was a lower proportion of people with high BP (22,556/140,100, 16.10% compared with 3386/14,690, 23.05%), a lower prevalence of diabetes (9339/180,290, 5.18% compared with 11,909/180,988, 6.58%), and a smaller proportion of people who smoked (22,470/180,481, 12.45% compared with 2149/15,394, 13.96%) in the SiSU health station group. There was also a lower proportion of people with BMI in the overweight to obese range (83,055/149,676, 55.49% compared with 92,651/149,727, 61.88%), with the average BMI almost 2 kg/m^2 lower for SiSU health station users (25.56 kg/m^2, SD 5.88 kg/m^2) compared with NHS participants (27.29 kg/m^2, SD 5.61 kg/m^2). Interestingly, however, Figures 1 and 2 demonstrate that the shape of the distribution of BMI for SiSU health station users and NHS participants is almost identical, with both distributions demonstrating right skewness and similar variability.

Table 2. Comparison of the health characteristics of SiSU health station users and National Health Survey participants.

Variables	SiSU health station users (unweighted), n (%)	SiSU health station users (weighted), n (%)	NHS[a] participants, n (%)	NHS participants vs weighted SiSU health station users, OR[b] (95% CI)[c]	*P* value[d]
High BP[e](mm Hg)					
Yes	22,556 (16.10)	26,905 (19.16)	3386 (23.05)	1.26 (1.21-1.31)	<.001
No	117,576 (83.90)	113,227 (80.84)	11,302 (76.95)	N/A[f]	N/A
BMI (kg/m^2) status					
Overweight/obese	83,055 (55.49)	92,651 (61.88)	9455 (62.07)	1.01 (0.97-1.04)	.66
Low to normal	66,624 (44.51)	67,245 (38.12)	5,779 (37.93)	N/A	N/A
Diabetes status					
Diabetic	9339 (5.18)	11,909 (6.58)	969 (6.30)	0.95 (0.89-1.02)	.17
Nondiabetic	171,103 (94.82)	168,533 (93.42)	14,424 (93.70)	N/A	N/A
Smoking status					
Nonsmoker	157,972 (87.55)	156,443 (86.70)	13,244 (86.04)	1.06 (1.01-1.11)	.02
Smoker	22,470 (12.45)	23,999 (13.30)	2149 (13.96)	N/A	N/A

[a] NHS: National Health Survey.

[b] OR: odds ratio.

[c] Relative odds of exposure in weighted SiSU users compared with NHS participants.

[d] *P* value was obtained from binary logistic regression.

[e] BP: blood pressure.

[f] N/A: not applicable.

Figure 1. BMI distribution (kg/m2) of SiSU Health Station users, unweighted.

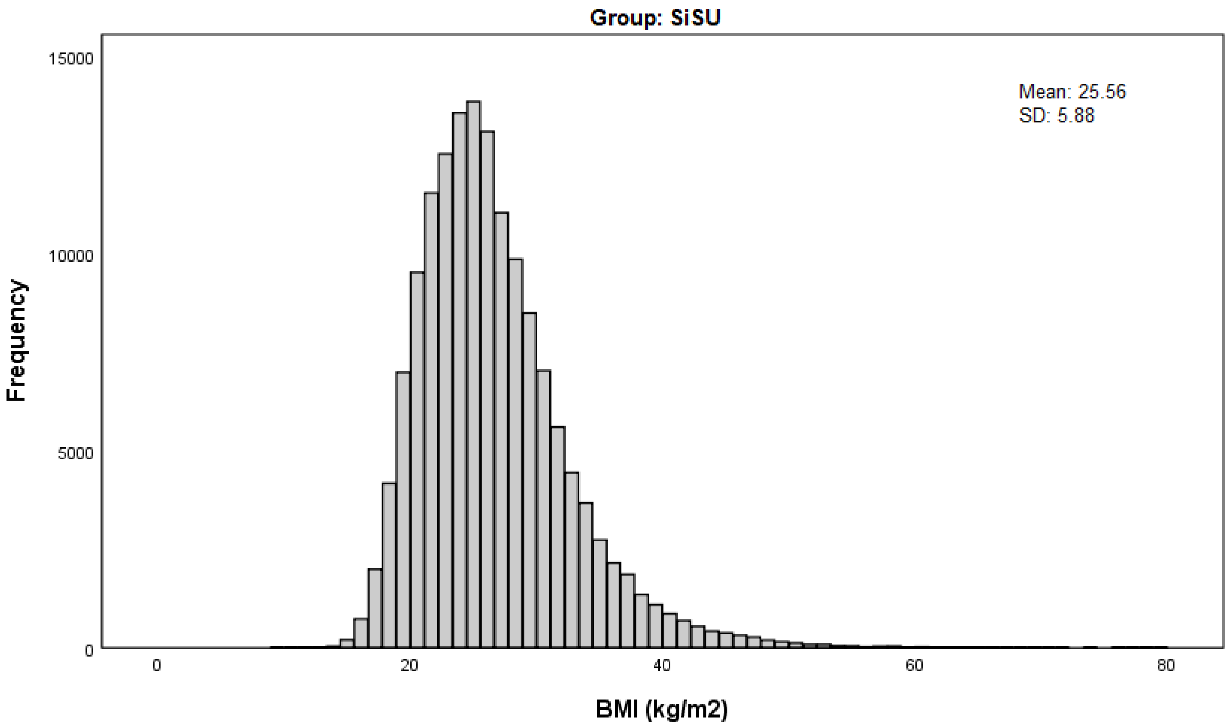
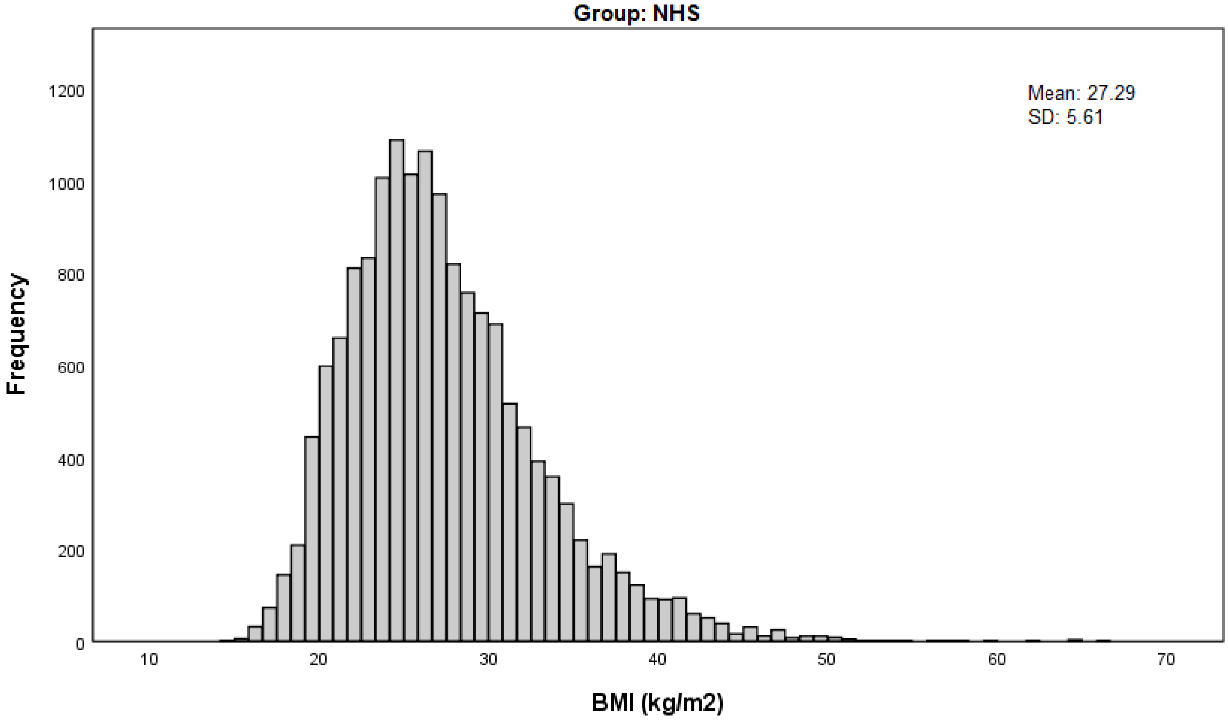


Figure 2. BMI distribution (kg/m2) of National Health Survey participants.



Given the demonstrated difference in demographics between the 2 groups, weights were applied to the SiSU health station data to account for these disparities, using the demographic variables age, gender, SES, and state to calculate the weights. The comparison of health characteristics was repeated using these weighted data to determine if SiSU health station users are healthier than the general population when these demographic differences are taken into consideration. ORs were calculated for each health characteristic to assess the magnitude of the difference between the weighted health characteristics of SiSU participants and those of the NHS participants.

Following the weighting of the SiSU health station data, the high BP and smoking variable distributions both demonstrated a significant difference with NHS participants. The proportion of SiSU health station users with high BP remained smaller compared with that of NHS participants (26,905/140,423, 19.16% compared with 3386/14,690, 23.05%), with NHS participants 1.26 times more likely to have high BP compared with SiSU users (*P*<.001). With regard to smoking status, the weighting of the SiSU data did result in a slight increase in the proportion of smokers; however, a smaller proportion of smokers remained in the SiSU group (23,999/180,444, 13.30% compared with 2149/15,394, 13.96%). Although this difference in the proportions was significant, the OR for this relationship was close to 1 (OR 1.06; *P*=.02), indicating that the difference is negligible.

Comparison of the BMI and diabetes status of SiSU health station users and NHS participants demonstrated no significant difference between proportions or an OR close to 1. The difference in the proportion of people with BMI in the overweight to obese range was not significant (92,651/149,727, 61.88% compared with 9455/15,233, 62.07%; OR 1.01; *P*=.66); although there was now a larger proportion of SiSU users with diabetes compared with the NHS participants (11,909/180,988, 6.58% compared with 969/15,381, 6.30%), this relationship was not significant (OR 0.95; *P*=.17).

Investigation of the Factors That Predict the Repeat Utilization of SiSU Wellness Health Check Stations

Binary logistic regression was used to identify demographic and health-related factors that are associated with repeat utilization of the SiSU health stations. Suspected shared accounts were removed from the analysis, resulting in a sample size of 179,467. Repeat users comprised 4.68% (8441/179,467) of all SiSU health station users in the sample, accounting for a total of 27,522 health checks, with a mean of 3.26 (SD 3.44) checks per person. The average time between baseline and final check for the users in this cohort was 77.60 days, with a minimum of 1 day and a maximum of 283 days between baseline and final check. Overall, 50.84% of users undertook their final check less than 90 days after their baseline check.

Overall, the model was found to be statistically significant (χ^2_{17}=483.5; *P*<.001), but only accounted for 1.00% of the total variance in the dependent variable, repeat user status (Nagelkerke R Square=0.011). Age was found to be significantly associated with being a repeat user, with the odds of being a repeat user decreasing with age; however, this effect size was small (OR 0.992; *P*<.001). Females were 1.371 times more likely to be repeat users (OR 1.371; *P*<.001), whereas those who resided in SEIFA quintiles 4 and 5 were significantly more likely to be repeat users compared with those residing in quintile 1 (OR 1.243; *P*<.001 and OR 1.151; *P*<.001, respectively). Only those residing in the Australian Capital Territory, Victoria, and South Australia had significantly lower odds of being repeat users compared with those residing in New South Wales (Table 3).

In terms of health characteristics, BMI, high BP, and smoking status were all significantly associated with being a repeat user. As BMI increased, the odds of being a repeat user increased; however, the effect size was small (OR 1.020; *P*<.001). Those with high BP were significantly more likely to be repeat users (OR 1.151; *P*<.001), whereas smokers were significantly less likely to be repeat users (OR 0.773; *P*<.001). There was no significant association between diabetes status and being a repeat user (Table 3).

Table 3. Binomial logistic regression analysis with the dependent variable: repeat user status.

Independent variable	OR[a] (95% CI)	*P* value[b]
Age (years)	0.992 (0.991-0.994)	<.001
Gender (reference=male)		
Female	1.371 (1.305-1.441)	<.001
SEIFA[c] (reference=quintile 1)		
Quintile 2	0.932 (0.837-1.037)	.12
Quintile 3	1.087 (0.993-1.190)	.07
Quintile 4	1.243 (1.146-1.349)	<.001
Quintile 5	1.151 (1.070-1.239)	<.001
State (reference=NSW[d])		
Australian Capital Territory	0.696 (0.564-0.859)	.01
Victoria	0.926 (0.859-0.999)	.048
Queensland	0.968 (0.907-1.032)	.32
South Australia	0.844 (0.751-0.949)	.01
Western Australia	0.980 (0.892-1.076)	.67
Tasmania	0.992 (0.814-1.209)	.94
Northern Territory	0.609 (0.248-1.494)	.28
BMI (kg/m^2)	1.020 (1.015-1.024)	<.001
High BP[e] (mm Hg; reference=no)		
Yes	1.151 (1.070-1.239)	<.001
Smoking status (reference=nonsmoker)		
Smoker	0.773 (0.718-0.831)	<.001
Diabetes status (reference=diabetic)		
Diabetic	0.927 (0.823-1.045)	.22

[a] OR: odds ratio.

[b] *P* value was obtained from binomial logistic regression.

[c] SEIFA: socioeconomic indexes for areas.

[d] NSW: New South Wales.

[e] BP: blood pressure.

Identify If the Health Status of Repeat Users of SiSU Health Stations in Australia Improved From Baseline to Final Check

A Wilcoxon signed rank test and McNemar tests were performed on the baseline and final health scores for the continuous and binary health characteristic variables, respectively. At baseline, the mean BMI was 26.37 kg/m^2 (SD 7.43 kg/m^2), decreasing to 25.40 kg/m^2 (SD 8.06 kg/m^2) at the final check. Results of the Wilcoxon signed rank test indicated that this change in mean BMI scores was significant (Z=−14.24; *P*<.001; Table 4).

The proportion of people with high BP decreased from baseline (5768/6848, 15.77%) to final check (885/6850, 12.92%). The results of the McNemar test confirmed that this change in proportions was significant (χ^2_1=38.2; *P*<.001). Of the 1080 users with high BP at baseline, 590 (54.63%) did not have high BP at their final check. Conversely, of the 5768 users who did not have high BP at baseline, 395 (6.85%) had high BP at their final check (Table 5).

The proportion of smokers was also found to decrease from baseline (1005/8438, 11.91%) to final check (854/8430, 10.13%). This decrease was found to be significant (χ^2_1=48.4; *P*<.001). Of the 1005 smokers at baseline, 308 (30.65%) were not smokers at their final check, whereas 2.11% (157/7436) of the nonsmokers at baseline reported smoking at their final check (Table 5).

Table 4. BMI of repeat SiSU users at baseline compared with final check.

Variable	Baseline, mean (SD)	Final check, mean (SD)	Test statistic (Z)	*P* value
BMI (kg/m^2)	26.37 (7.43)	25.40 (8.06)	−14.24	<.001

Table 5. Blood pressure and smoking status of SiSU users at baseline compared with final check

Variables	Baseline, n (%)	Final check, n (%)	Chi-square value (*df*=1)	*P* value
High blood pressure (mm Hg)				
No	5768 (84.23)	5963 (87.08)	38.2	<.001
Yes	1080 (15.77)	885 (12.92)	N/A[a]	N/A
Smoking status				
Nonsmoker	7436 (88.09)	7578 (89.87)	48.4	<.001
Smoker	1005 (11.91)	854 (10.13)	N/A	N/A

[a] N/A: not applicable.

Finally, linear models were used to determine if the number of health checks a user undertakes was significantly related to their final health scores. The baseline scores were controlled in these models.

Due to the skewed distribution of BMI in the final check variables, a generalized linear model with gamma distribution and log link was used to assess the relationship between final BMI measurements and number of health checks. When controlling for baseline measures, the exponentiated coefficient (Exp[b]=0.999; *P*=.23) indicates that BMI at the final check was not significantly associated with the number of health checks (Table 6).

Binary logistic regression demonstrated that an increasing number of health checks were significantly associated with a decreasing likelihood of being a smoker at the final check, although this effect was small (OR 0.959; *P*<.048). There was no significant relationship between the number of health checks and high BP at the final check (OR 0.985; *P*=.14; Table 7).

Table 6. Generalized linear model explaining the effect of number of health checks on BMI at final check.

Parameter	Exp (b)[b] (SE)	*P* value
Number of health checks	0.999 (1.001)	.23
BMI at initial check	1.021 (1.000)	<.001
Gender (reference=female)	1.017 (1.003)	<.001
SEIFA[a] quintile (reference=quintile 5)		<.001
Quintile 1	1.046 (1.005)	
Quintile 2	1.035 (1.006)	
Quintile 3	1.033 (1.005)	
Quintile 4	1.019 (1.004)	
Age	1.001 (1.001)	<.001

[a] SEIFA: socioeconomic indexes for areas.

[b] Exp(b): exponentiated coefficient.

Table 7. Binary logistic regression models explaining the effect of number of health checks on smoking status and high blood pressure status at final check.

Parameter	Smoking status at final check (reference=nonsmoker)		High BP[a] (mm Hg) at final check (reference=no)	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value
Number of health checks	0.959 (0.921-1.000)	.048	0.985 (0.964-1.005)	.14
High BP at initial check	N/A[c]	N/A	0.113 (0.096-0.132)	<.001
Smoking status at initial check	0.010 (0.008-0.012)	<.001	N/A[c]	N/A
Gender (reference=female)	1.074 (0.870-1.324)	.51	1.343 (1.145-1.574)	<.001
SEIFA[d] quintile (reference=quintile 5)				
Quintile 1	1.696 (1.277-2.252)	<.001	1.348 (1.075-1.691)	.01
Quintile 2	1.296 (0.888-1.892)	.18	1.220 (0.907-1.640)	.19
Quintile 3	1.403 (1.007-1.953)	.045	1.500 (1.183-1.901)	.001
Quintile 4	1.001 (0.753-1.331)	.99	1.145 (0.919-1.425)	.23
Age	0.991 (0.984-0.998)	.02	1.028 (1.023-1.033)	<.001

[a] BP: blood pressure.
[b] OR: odds ratio.
[c] N/A: Not applicable.
[d] SEIFA: socioeconomic indexes for areas.

Discussion

Principal Findings

At 180,000 health checks, the SiSU wellness dataset is one of the largest datasets ever provided for research, which has been generated by interactive health stations that measure biometric indicators. The scale of this dataset is clearly significant, and this study and the SiSU wellness dataset provide a valuable foundation for extensive investigation into the benefits of health stations in terms of their health monitoring and health promotion capabilities.

This study builds on previous research in the health technology field by providing further insight into the factors that are associated with health kiosk utilization and the potential health benefits of using health kiosks to self-monitor health. Findings indicate that demographics, including gender, age, and SES, were associated with both utilization and repeat utilization of the SiSU health stations, with females, younger people, and those of higher SES using the SiSU health stations at higher rates and more likely to be repeat users. A relationship between health characteristics and repeat utilization of the health stations has also been demonstrated; both higher BMI and high BP at baseline increased the odds of being a repeat user, and smokers were less likely to be repeat users.

The results are consistent with findings regarding the influence of age on the utilization of health technologies, which suggest that younger people use these technologies at higher rates [23-25]. Despite the suggestion by some authors that health kiosks may play a role in reducing this age bias in the use of technology by reducing age-related barriers [29-31], this does not appear to be supported by the results of our study.

Previous studies investigating the influence of gender on health kiosk utilization have demonstrated mixed results. Consistent with our findings, 1 study found that females access health kiosks at higher rates [29], whereas another study found that the relationship between gender and kiosk use is dependent on other demographic variables such as country of birth and SES [32], and other studies have found no relationship between gender and kiosk use [30,33]. The inconsistency of these results is potentially due to differences in the location of the health kiosks (eg, retail environment and hospital setting), which has been found to influence utilization rates [31]; further research could aim to investigate these effects. It is worth noting, however, that although a smaller proportion of SiSU health station users are males, the sheer volume of males undertaking health checks using the SiSU health stations provides a unique opportunity for engaging this demographic, who have been found to access traditional preventative health services at lower rates than females [43,44].

When considering the gender distribution of Priceline Pharmacies' customer base in conjunction with that of SiSU health stations users, evidence emerges that suggests that the SiSU health stations are highly effective in engaging the male demographic, with 44.13% (79,628/180,442) of SiSU health station users being males. Evidence to support the hypothesis that the SiSU health station is particularly attractive to males in retail environments is further supported by unpublished data from a 6-store retail health station pilot that SiSU wellness undertook with a major Australian Supermarket chain, where males accounted for 54.29% (50,143/92,345) of the checks recorded. In 5 of the 6 supermarkets, males contributed the clear majority of checks, and in the only store where females were the majority, the spread between genders was just 1.40% (Personal Communication by Patrick J Hannebery, May 29, 2019).

The findings regarding the SES of SiSU health station users support previous research that demonstrate that higher SES levels are associated with higher utilization rates of health kiosks [30,31] and health technologies in general [23,25,28]. The poorer rates of utilization in lower SES populations may be the result of a combination of factors, including lower health literacy [45,46]; access barriers such as cost, time, and transport [46,47]; and attitudes toward health care and health care providers [47,48]. Again, location of the health kiosks may play a role in the lower rate of utilization by those from low SES areas.

Our initial descriptive analysis of the health characteristics of SiSU health station users found that SiSU users are generally healthier than the NHS participants on the 4 health characteristics investigated in this study. This supports the findings of previous research, which found that people with chronic disease are less likely to engage in self-monitoring. Such research suggests that poor health is associated with decreased self-efficacy and confidence in health monitoring and improvement, sometimes leading to decreased utilization and adoption of monitoring and prevention [26], which may in part explain these findings. The results of the weighted analysis, however, indicate that age, gender, SES, and place of residence (state) play a large role in the difference in the disease status between SiSU users and the general population, with BP being the only indicator with a substantially large difference between SiSU users and the NHS participants once data were weighted to address the disparity in demographics.

Our investigation of the factors associated with repeat utilization, however, demonstrated that people with high BP are more likely to be repeat users of the health stations and that as BMI increases, the odds of being a repeat user increases. This is an interesting finding in that SiSU users were found to have more favorable outcomes on the 4 health indicators under investigation, whereas repeat users were more likely to have high BP, diabetes, and higher BMIs. These findings provide insight into the differences between once-off utilization of health technologies as opposed to sustained use. The placement of the majority of SiSU health stations in pharmacies may play a role in this relationship, given that those with health issues may be more likely to go to pharmacies on a regular basis.

Regarding repeat utilization of SiSU health stations, our results are somewhat consistent with previous research that demonstrate that people with chronic disease are more likely to use technology to self-monitor their health in a sustained manner [24-27]. For example, higher BMI has been found to be associated with higher frequency of use of mobile health apps [25], whereas those with hypertension have previously demonstrated a higher willingness to consistently self-monitor their health [26]. The difference in the health characteristics of repeat users compared with one-time users is interesting. It has been suggested by some authors that sustained use of self-monitoring may be driven by health-related goals such as weight loss [24,49], which may partially explain these findings.

This study has also provided an initial analysis of the health outcomes of health kiosk users, which to date, have not been established. In doing so, it has paved the way for further investigation into the benefits of such technologies for health monitoring by identifying potential areas for further research.

The findings of our study demonstrate positive initial results in terms of the change in health status between baseline and final health checks for repeat users, although the number of health checks a person undertakes does not appear to influence outcomes at the final check. In terms of the comparison of baseline and final checks, these results are somewhat consistent with much of the research that has been conducted regarding the benefits of various self-monitoring health technologies [6,50,51]; however, they are somewhat inconsistent with the finding that the number of health checks does not impact health outcomes at the final check. For example, previous research has found that the adherent use of digital trackers is associated with weight loss and increased physical activity [6], whereas other studies have found that regular monitoring of BP leads to reductions in BP when combined with other interventions [50,51].

Although the suggestion that self-monitoring is effective when combined with other health interventions is interesting, unfortunately it is difficult from the SiSU data to determine the method through which improvements were achieved (eg, counseling, medication, and physical activity) and, if so, whether the use of SiSU health stations influenced the uptake of these methods. Due to the dearth of research regarding the benefits of technology for self-monitoring, there is also limited information available regarding the possible mechanisms behind the behavior changes that result in the health improvements observed in our study and other studies. Research into the effect of self-monitoring in general on health status suggests that mechanisms may include empowerment or self-efficacy [8,13], self-actualization and self-esteem, as well as greater sensitivity and awareness [6]. With regard to health technologies, there is also some evidence that the aggregation of health data in 1 place provides a more complete picture of health status, enabling more holistic behavior change [14]. Again, these findings present further research opportunities with regard to SiSU health station users.

Finally, there is also some evidence that the benefits of digital health monitoring are more pronounced for people with more serious health problems [6], again demonstrating the potential for further investigation of the SiSU data with regard to this finding.

It is worth noting that the results of all analyses undertaken have the potential to be influenced by the location of the SiSU health stations, that is, pharmacy, retail shopping center, gym, or corporate setting, rural versus metropolitan area, and other location variables. There is potential that the differences in demographics and health status between the SiSU wellness users and NHS participants may be partially explained by the placement of stations across the country, with more health check stations in New South Wales than other states. Repeat utilization and frequency of use of the SiSU health stations are also likely to be influenced by these factors. Future research should aim to investigate the effect of location on the demographic representatives and health characteristics of all users, the factors associated with repeat utilization, and the changes in the health characteristics of repeat users.

Overall, the results of this study are promising in terms of the potential for utilization of SiSU health stations as an effective catalyst for change and a means of empowering consumers to take ownership and achieve improvements in their health status. The considerably large sample size, digitization, and representativeness of health status measures in the data demonstrate the unique value of the SiSU health station data as a cost-effective method of monitoring population health data over time and provides a valuable data source for health workers, population health professionals, policy makers, and researchers alike.

Limitations

A number of potential limitations of this study have been identified. First, the range of demographic data available in both the SiSU wellness and ABS datasets limit the investigation of the representativeness of the SiSU data. It is possible that there may be underlying differences between the SiSU wellness sample and the general population that cannot be identified through this study. The limited number of demographic variables available for analysis and comparison also mean that the weighting of the SiSU wellness data is restricted to variables available in both the SiSU wellness dataset and the NHS dataset. Furthermore, only a small percentage of the variance in the repeat status of users could be explained by the available variables; additional variables would allow more robust models to be developed to investigate the relationship between demographics, health status, and repeat users more thoroughly.

It must also be acknowledged that the SiSU wellness data and the ABS census/NHS data are not mutually exclusive; in that it is possible that subjects in the SiSU wellness dataset could also be participants in the NHS. Due to this, and as there is no sampling frame to directly compare users and nonusers of the SiSU health stations, interpretation of the results of the comparison of SiSU users and NHS participants should be treated with caution. There are also limitations associated with comparing data from different sources. Although every effort has been made to ensure that the variables from the SiSU health station and NHS data are directly comparable, by collapsing response categories where necessary and using actual measurements to compute categories for variables such as BMI and BP, we were unable to control for potential differences in measurement techniques.

A further limitation is associated with the study design, as it is difficult to determine from the results if the utilization of SiSU health stations has led, directly or indirectly, to the improvements in health characteristics observed in this study. Data available through the SiSU health stations include Prochaska and DiClemente’s *Stages of Change* assessment data, and it is recommended that further studies investigating the relationship between SiSU health station utilization and health outcomes incorporate the relationship between these responses, repeat utilization of health stations, and health outcomes. Additional considerations for further investigation include the presence or absence of other potential drivers of change, including contact with health practitioners and healthy lifestyle programs, and potentially the inclusion of a control group to identify if these improvements are only observed in SiSU users. It must also be noted that the statistical methodology used for this particular research question did not allow for the inclusion of covariates; future research should aim to develop more complex models that involve a wider range of explanatory variables. Finally, the time elapsed between baseline and final check was not considered in these models and could be included in future to identify if changes to health status are more pronounced and/or sustained over time.

For these reasons, although contributing to the small body of research regarding the factors associated with and benefits of health kiosks for self-monitoring health, care should be taken in generalizing these results. Further research should build upon these initial investigations to address some of the identified limitations.

Conclusions

The findings of this study support much of the previous research regarding the relationship between demographics, health status, and uptake of self-monitoring for health, and, in particular, provide valuable insights regarding the health benefits of health kiosks for the self-monitoring of health. In general, users of SiSU health stations differ from the general population in terms of demographics and are likely to be healthier than the general population. However, this difference in health status appears to be minimal when differences in demographics are taken into account. The results of the study are promising in terms of the potential benefits of using SiSU wellness stations to monitor health; however, there are opportunities for further research into the factors relating to and mechanisms behind these benefits.

Acknowledgments

The authors would like to acknowledge Dr Noel Duncan, PhD and founder of SiSU Health Group, for providing the data to enable this research.

Conflicts of Interest

None declared.

References

1. Spring B, Gotsis M, Paiva A, Spruijt-Metz D. Healthy apps: mobile devices for continuous monitoring and intervention. IEEE Pulse 2013;4(6):34-40 [FREE Full text] [doi: 10.1109/MPUL.2013.2279620] [Medline: 24233190]

2. Kim JY, Wineinger NE, Taitel M, Radin JM, Akinbosoye O, Jiang J, et al. Self-monitoring utilization patterns among individuals in an incentivized program for healthy behaviors. J Med Internet Res 2016 Nov 17;18(11):e292 [FREE Full text] [doi: 10.2196/jmir.6371] [Medline: 27856407]
3. McBain H, Shipley M, Newman S. The impact of self-monitoring in chronic illness on healthcare utilisation: a systematic review of reviews. BMC Health Serv Res 2015 Dec 18;15:565 [FREE Full text] [doi: 10.1186/s12913-015-1221-5] [Medline: 26684011]
4. Tanenbaum ML, Bhatt HB, Thomas VA, Wing RR. Use of self-monitoring tools in a clinic sample of adults with type 2 diabetes. Transl Behav Med 2017 Jun;7(2):358-363 [FREE Full text] [doi: 10.1007/s13142-016-0418-4] [Medline: 27270913]
5. Orji R, Lomotey R, Oyibo K, Orji F, Blustein J, Shahid S. Tracking feels oppressive and 'punishy': exploring the costs and benefits of self-monitoring for health and wellness. Digit Health 2018;4:2055207618797554 [FREE Full text] [doi: 10.1177/2055207618797554] [Medline: 30202544]
6. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 9;13:77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
7. Laranjo L, Lau A, Oldenburg B, Gabarron E, O'Neill A, Chan S, et al. Sax Institute. 2015. mHealth Technologies for Chronic Disease Prevention and Management URL: https://www.saxinstitute.org.au/wp-content/uploads/mHealth-Technologies-for-Chronic-Disease-Prevention-and-Management-1.pdf [accessed 2020-05-04]
8. Karapanos E, Gouveia R, Hassenzahl M, Forlizzi J. Wellbeing in the making: peoples' experiences with wearable activity trackers. Psychol Well Being 2016;6:4 [FREE Full text] [doi: 10.1186/s13612-016-0042-6] [Medline: 27376017]
9. Pourzanjani A, Quisel T, Foschini L. Adherent use of digital health trackers is associated with weight loss. PLoS One 2016;11(4):e0152504 [FREE Full text] [doi: 10.1371/journal.pone.0152504] [Medline: 27049859]
10. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
11. Vaes AW, Cheung A, Atakhorrami M, Groenen MT, Amft O, Franssen FM, et al. Effect of 'activity monitor-based' counseling on physical activity and health-related outcomes in patients with chronic diseases: a systematic review and meta-analysis. Ann Med 2013 Sep;45(5-6):397-412. [doi: 10.3109/07853890.2013.810891] [Medline: 23952917]
12. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
13. Lupton D. Quantifying the body: monitoring and measuring health in the age of mhealth technologies. Critical Public Health 2013 Dec;23(4):393-403. [doi: 10.1080/09581596.2013.794931]
14. Gay V, Leijdekkers P. Bringing health and fitness data together for connected health care: mobile apps as enablers of interoperability. J Med Internet Res 2015 Nov 18;17(11):e260 [FREE Full text] [doi: 10.2196/jmir.5094] [Medline: 26581920]
15. Shammas L, Zentek T, von Haaren B, Schlesinger S, Hey S, Rashid A. Home-based system for physical activity monitoring in patients with multiple sclerosis (pilot study). Biomed Eng Online 2014 Feb 6;13:10 [FREE Full text] [doi: 10.1186/1475-925X-13-10] [Medline: 24502230]
16. Richman W, Kiesler S, Weisband S, Drasgow F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. J Appl Psychol 1999 Oct;84(5):754-775. [doi: 10.1037/0021-9010.84.5.754]
17. Huerta JM, Tormo MJ, Egea-Caparrós JM, Ortolá-Devesa JB, Navarro C. Accuracy of self-reported diabetes, hypertension and hyperlipidemia in the adult Spanish population. DINO study findings. Rev Esp Cardiol 2009 Feb;62(2):143-152 [FREE Full text] [doi: 10.1016/s1885-5857(09)71532-4] [Medline: 19232187]
18. Taylor A. University of Adelaide. 2006. Representative Population Health Surveys: Improving Public Health Through Rigour, Diversity of Methods and Collaboration URL: https://digital.library.adelaide.edu.au/dspace/bitstream/2440/37891/9/01front.pdf [accessed 2020-05-04]
19. Stergiou G, Mengden T, Padfield PL, Parati G, O'Brien E, Working Group on Blood Pressure Monitoring of the European Society of Hypertension. Self monitoring of blood pressure at home. Br Med J 2004 Oct 16;329(7471):870-871 [FREE Full text] [doi: 10.1136/bmj.329.7471.870] [Medline: 15485940]
20. Leary MR, Kowalski RM. Impression management: a literature review and two-component model. Psychol Bull 1990;107(1):34-47. [doi: 10.1037/0033-2909.107.1.34]
21. Chun H, Kim I, Min K. Accuracy of self-reported hypertension, diabetes, and hypercholesterolemia: analysis of a representative sample of Korean older adults. Osong Public Health Res Perspect 2016 Apr;7(2):108-115 [FREE Full text] [doi: 10.1016/j.phrp.2015.12.002] [Medline: 27169009]
22. Pierin AM, Mano G, Souza V, de Lima JC, Ortega KC, Caetano E, et al. Blood pressure measurement taken by patients or nurses avoids observer's influence. Am J Hypertens 2001 Nov;14(11):A44-A50 [FREE Full text] [doi: 10.1016/s0895-7061(01)01521-7]
23. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]

24. Hermsen S, Moons J, Kerkhof P, Wiekens C, de Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
25. Krebs P, Duncan DT. Health app use among us mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
26. Huygens MW, Swinkels IC, de Jong JD, Heijmans MJ, Friele RD, van Schayck OC, et al. Self-monitoring of health data by patients with a chronic disease: does disease controllability matter? BMC Fam Pract 2017 Mar 20;18(1):40 [FREE Full text] [doi: 10.1186/s12875-017-0615-3] [Medline: 28320330]
27. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among us mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
28. Bandura A. The primacy of self-regulation in health promotion. Appl Psychol 2005 Apr;54(2):245-254. [doi: 10.1111/j.1464-0597.2005.00208.x]
29. Pendleton BF, Schrop SL, Ritter C, Kinion ES, McCord G, Cray JJ, et al. Underserved patients' choice of kiosk-based preventive health information. Fam Med 2010;42(7):488-495 [FREE Full text] [Medline: 20628922]
30. Jones R. The role of health kiosks in 2009: literature and informant review. Int J Environ Res Public Health 2009 Jun;6(6):1818-1855 [FREE Full text] [doi: 10.3390/ijerph6061818] [Medline: 19578463]
31. Nicholas D, Huntington P, Williams P. Health kiosk use: a national comparative study. Aslib Proc 2001 May;53(4):130-140. [doi: 10.1108/EUM0000000007048]
32. Nicholas D, Huntington P, Williams P, Chahal P. Determinants of health kiosk use and usefulness: case study of a kiosk which serves a multi-cultural population. Libri 2001 Jan;51(2):102-113. [doi: 10.1515/LIBR.2001.102]
33. Chan YF, Nagurka R, Bentley S, Ordonez E, Sproule W. Medical utilization of kiosks in the delivery of patient education: a systematic review. Health Promot Perspect 2014;4(1):1-8 [FREE Full text] [doi: 10.5681/hpp.2014.001] [Medline: 25097831]
34. SiSU Wellness. 2018. Measure, Track, Improve URL: https://www.sisuwellness.com.au/sisu-station [accessed 2020-02-14]
35. Australian Bureau of Statistics. 2015. 4364.0.55.001 - National Health Survey: First Results, 2017-18 URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/4364.0.55.001 [accessed 2018-06-11]
36. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res 2013 Jun;22(3):278-295. [doi: 10.1177/0962280210395740] [Medline: 21220355]
37. Pourhoseingholi MA, Baghestani A, Vahedi M. How to control confounding effects by statistical analysis. Gastroenterol Hepatol Bed Bench 2012;5(2):79-83 [FREE Full text] [Medline: 24834204]
38. Sperandei S. Understanding logistic regression analysis. Biochem Med (Zagreb) 2014;24(1):12-18 [FREE Full text] [doi: 10.11613/BM.2014.003] [Medline: 24627710]
39. Lachenbruch PA, Lynch CJ. Assessing screening tests: extensions of McNemar's test. Stat Med 1998 Oct 15;17(19):2207-2217. [doi: 10.1002/(sici)1097-0258(19981015)17:19<2207::aid-sim920>3.0.co;2-y] [Medline: 9802179]
40. Rosner B, Glynn RJ, Lee MT. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 2006 Mar;62(1):185-192. [doi: 10.1111/j.1541-0420.2005.00389.x] [Medline: 16542245]
41. Dodd S, Bassi A, Bodger K, Williamson P. A comparison of multivariable regression models to analyse cost data. J Eval Clin Pract 2006 Feb;12(1):76-86. [doi: 10.1111/j.1365-2753.2006.00610.x] [Medline: 16422782]
42. Green R. Campaign Brief. 2016. Priceline Pharmacy Targets Aussie Females in ‘100% Woman’ Campaign via Ogilvy, Melbourne URL: https://campaignbrief.com/priceline-pharmacy-targets-aus/ [accessed 2020-05-17]
43. Keene J, Li X. Age and gender differences in health service utilization. J Public Health (Oxf) 2005 Mar;27(1):74-79. [doi: 10.1093/pubmed/fdh208] [Medline: 15637107]
44. Redondo-Sendino A, Guallar-Castillón P, Banegas JR, Rodríguez-Artalejo F. Gender differences in the utilization of health-care services among the older adult population of Spain. BMC Public Health 2006 Jun 16;6:155 [FREE Full text] [doi: 10.1186/1471-2458-6-155] [Medline: 16780576]
45. Akinyemiju TF. Socio-economic and health access determinants of breast and cervical cancer screening in low-income countries: analysis of the world health survey. PLoS One 2012;7(11):e48834 [FREE Full text] [doi: 10.1371/journal.pone.0048834] [Medline: 23155413]
46. Brunner-Ziegler S, Rieder A, Stein KV, Koppensteiner R, Hoffmann K, Dorner TE. Predictors of participation in preventive health examinations in Austria. BMC Public Health 2013 Dec 5;13:1138 [FREE Full text] [doi: 10.1186/1471-2458-13-1138] [Medline: 24308610]
47. Zhang J. QUT ePrints. 2007. eprints.qut.edu.au URL: https://eprints.qut.edu.au/16532/1/Jianzhen_Zhang_Thesis.pdf [accessed 2018-10-24]
48. Wilf-Miron R, Peled R, Yaari E, Vainer A, Porath A, Kokia E. The association between socio-demographic characteristics and adherence to breast and colorectal cancer screening: analysis of large sub populations. BMC Cancer 2011 Aug 25;11:376 [FREE Full text] [doi: 10.1186/1471-2407-11-376] [Medline: 21867544]
49. Wang J, Sereika S, Chasens ER, Ewing LJ, Matthews JT, Burke LE. Effect of adherence to self-monitoring of diet and physical activity on weight loss in a technology-supported behavioral intervention. Patient Prefer Adher 2012;6:221-226 [FREE Full text] [doi: 10.2147/PPA.S28889] [Medline: 22536058]

50. Tucker KL, Sheppard JP, Stevens R, Bosworth HB, Bove A, Bray EP, et al. Self-monitoring of blood pressure in hypertension: a systematic review and individual patient data meta-analysis. PLoS Med 2017 Sep;14(9):e1002389 [FREE Full text] [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]
51. Bray EP, Holder R, Mant J, McManus RJ. Does self-monitoring reduce blood pressure? Meta-analysis with meta-regression of randomized controlled trials. Ann Med 2010 Jul;42(5):371-386. [doi: 10.3109/07853890.2010.489567] [Medline: 20504241]

Abbreviations

ABS: Australian Bureau of Statistics
BP: blood pressure
IPW: inverse probability weighting
IRSAD: index of relative advantage and disadvantages
NHS: National Health Survey
OR: odds ratio
SEIFA: socioeconomic indexes for areas
SES: socioeconomic status

Edited by G Eysenbach; submitted 09.06.19; peer-reviewed by C Fincham, R Nogueira-Arjona; comments to author 29.10.19; revised version received 20.11.19; accepted 15.12.19; published 26.06.20.

Please cite as:
Flitcroft L, Chen WS, Meyer D
The Demographic Representativeness and Health Outcomes of Digital Health Station Users: Longitudinal Study
J Med Internet Res 2020;22(6):e14977
URL: http://www.jmir.org/2020/6/e14977/
doi:10.2196/14977
PMID:32589150

Original Paper

Continuous Monitoring of Vital Signs in the General Ward Using Wearable Devices: Randomized Controlled Trial

Mariska Weenk[1], MD, PhD; Sebastian J Bredie[1], MD, PhD; Mats Koeneman[1]; Gijs Hesselink[1], PhD; Harry van Goor[1], MD, PhD; Tom H van de Belt[1], PhD

[1] Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Mariska Weenk, MD, PhD
Radboud University Medical Center
PO Box 9101
Nijmegen, 6500 HB
Netherlands
Phone: 31 621894189
Email: mariska.weenk@radboudumc.nl

Abstract

Background: Wearable devices can be used for continuous patient monitoring in the general ward, increasing patient safety. Little is known about the experiences and expectations of patients and health care professionals regarding continuous monitoring with these devices.

Objective: This study aimed to identify positive and negative effects as well as barriers and facilitators for the use of two wearable devices: ViSi Mobile (VM) and HealthPatch (HP).

Methods: In this randomized controlled trial, 90 patients admitted to the internal medicine and surgical wards of a university hospital in the Netherlands were randomly assigned to continuous vital sign monitoring using VM or HP and a control group. Users' experiences and expectations were addressed using semistructured interviews. Nurses, physician assistants, and medical doctors were also interviewed. Interviews were analyzed using thematic content analysis. Psychological distress was assessed using the State Trait Anxiety Inventory and the Pain Catastrophizing Scale. The System Usability Scale was used to assess the usability of both devices.

Results: A total of 60 patients, 20 nurses, 3 physician assistants, and 6 medical doctors were interviewed. We identified 47 positive and 30 negative effects and 19 facilitators and 36 barriers for the use of VM and HP. Frequently mentioned topics included earlier identification of clinical deterioration, increased feelings of safety, and VM lines and electrodes. No differences related to psychological distress and usability were found between randomization groups or devices.

Conclusions: Both devices were well received by most patients and health care professionals, and the majority of them encouraged the idea of monitoring vital signs continuously in the general ward. This comprehensive overview of barriers and facilitators of using wireless devices may serve as a guide for future researchers, developers, and health care institutions that consider implementing continuous monitoring in the ward.

Trial Registration: Clinicaltrials.gov NCT02933307; http://clinicaltrials.gov/ct2/show/NCT02933307.

(J Med Internet Res 2020;22(6):e15471) doi:10.2196/15471

KEYWORDS

remote sensing technology; wireless technology; continuous monitoring; vital signs; wearable electronic devices; remote monitoring; digital health

Introduction

Background

Today's technology is increasingly influencing health care [1]. Numerous wearable devices such as patches, smart watches, and even tattoos exist that can register vital signs such as heart rate (HR), respiratory rate, oxygen saturation (SpO_2), and blood pressure (BP) [2-5]. These devices are increasingly accurate and reliable [2,6], smaller, and more user friendly than current hospital monitoring devices. This could facilitate patients' mobility and recovery during admission [7,8]. Moreover, the devices can result in improved health outcomes and can be used as a diagnostic tool in the identification of several diseases or clinical deterioration during admission [2,9-11].

Clinical Deterioration

Vital signs of patients in general wards are usually monitored periodically by nurses, primarily during daytime [12]. As a result, clinical deterioration in between two subsequent measurements may not always be detected and can result in unplanned admission to the intensive care unit (ICU), which is associated with longer hospital stay, increased mortality rate [13-15], and higher costs [16]. During night hours, when less medical personnel are available, clinical deterioration may remain undetected until the next morning [17]. With wearable devices, patients can be monitored more frequently or continuously. This results in additional information about a patient's health status, particularly during out of office hours when patients are less frequently seen by nurses [4]. By implementing continuous monitoring, clinical deterioration can be detected in an earlier phase, particularly as changes in vital signs are often present 8 to 24 hours before a life-threatening event occurs [18-22]. Additional benefits of wearable device–based continuous monitoring are a reduced workload in nurses [23], improved patient comfort because of fewer vital sign measurements [8,24], and safe patient transport between wards [25]. Besides positive effects of wearable devices, continuous monitoring can lead to false alarms that result in unnecessary additional diagnostic procedures and possible alarm fatigue in health care professionals [26,27].

Wearable Devices

Recently, ViSi Mobile (VM) and the HealthPatch (HP) were introduced to hospital care. These two wearable devices are approved by the Food and Drug Association for continuous vital sign monitoring and have shown to be as accurate as nurse measurements in admitted patients [6]. Several studies describing the opportunities of wearable devices including VM and HP were primarily focused on the accuracy of data [11]. For successful implementation in hospital wards, wearable devices for continuous monitoring of vital signs should be comfortable and user friendly for both patients and health care professionals. Besides, patients and health care professionals should be willing to use them and see the benefit of these wearable devices and of being monitored continuously. A complete overview of experiences and expectations of patients regarding continuous monitoring with wearable devices is lacking. Therefore, this study aimed to identify experiences of patients, relatives, nurses, physician assistants, and medical doctors about the use of VM and HP in daily practice for continuous monitoring of vital signs in the general ward.

Methods

Setting, Participants, and Sampling

This randomized controlled trial was conducted in a university hospital between April 2015 and August 2016. The objective of this study was to give an overview of the experiences and expectations regarding continuous monitoring with wearable devices by most important stakeholders. This design was chosen as a control group would give an insight into the current experiences of patients who were not yet influenced by the use of wearable devices. Besides patients, the target population consisted of nurses, physician assistants, and medical doctors who were involved in the care of the included patients. Surgical patients were included when they were scheduled for an elective abdominal surgical procedure. Patients were excluded and replaced when they were monitored for less than 24 hours. A sample size of 90 patients (45 surgical patients and 45 internal medicine patients) was estimated to be sufficient to obtain data saturation regarding interviews, based on our pilot study [6]. As there are no standards to calculate sample size for qualitative research [28], we focused on data saturation. This was defined as the moment when additional interviews would not result in new information or themes, which was discussed and decided by two experienced qualitative researchers (MW and TB). Patients' relatives were involved if they attended the interview. We aimed to interview all nurses, physician assistants, and medical doctors who were involved in the care of the included patients to obtain a complete overview of users' experiences and expectations. The institutional review board decided that formal approval was not required after they reviewed the study protocol extensively (Local Ethical Committee number 2015-1717). The study was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

Wearable Devices

VM (Sotera Wireless) is a patient monitoring system developed to enhance patient safety and early detection of clinical deterioration in a general ward. VM continuously measures 5-lead electrocardiography (ECG), HR, respiratory rate, SpO_2, BP, and skin temperature. It transmits all data wirelessly to a platform with Sotera's analytic software such as desktop PCs or tablet PCs from where health care professionals have a real-time insight into patients' vital sign data. VM consists of a wrist device with a touch screen display that shows vital signs and a thumb sensor that measures SpO_2 and BP. Five ECG cables and a chest sensor that measures skin temperature and respiratory rate are attached to the patient's chest. The battery in the wrist device has to be changed every 12 to 16 hours.

The HP (Vital Connect) is a small and lightweight disposable adhesive patch that consists of two ECG electrodes and a reusable module, which contains a sensor and a Bluetooth transmitter. It contains a battery that has a wear cycle of approximately 3 to 4 days. The patch continuously measures 1-lead ECG, HR, respiratory rate, HR variability, skin temperature, steps, and body posture [29]. The patch is attached to the patient's chest, from where it sends data via Bluetooth to a mobile device where patients can see their own vital signs. Data are transmitted to a secured Vital Connect cloud on the internet via Wi-Fi.

Study Procedures and Data Collection

Interviews

Patients in the surgical and internal medicine wards provided written informed consent after being informed about the study protocol. All interviewed nurses, physician assistants, and medical doctors also signed the informed consent form. Patients were randomly assigned to (1) VM, (2) HP, or (3) control group (no device; 1:1:1). This was done to equalize individual factors between groups and minimize bias. The control group only received the regular nurse measurements. They were interviewed about their current experiences and their expectations of continuous monitoring, without being influenced by wearing a device. In the internal medicine ward, patients were randomized immediately after signing the informed consent form. Surgical patients signed the informed consent form before an elective surgical procedure and were randomized after surgery on arrival in the ward. Vital signs were continuously measured for 2 to 3 days in the VM and HP groups. Regular vital sign measurements (three times a day) by nurses continued according to the hospital protocol for all patients.

At the end of the study, patients and their relatives were interviewed face-to-face for approximately 45 min by one trained investigator. Nurses, physician assistants, and medical doctors who were involved in the care of the included patients were also interviewed. For each semistructured interview, an interview guide was used that consisted of predetermined themes based on the model for implementation by Grol and Wensing [30]. We added themes identified in a recent pilot study about monitoring with similar wearable devices [6]. Themes concerned attitude toward continuous monitoring and the wearable devices, experiences with both wearables in clinical practice, future expectations of the devices, and perception on changes in clinical care using the devices. Questions focused on, for example, feelings of safety, users' experiences with the devices, expected effect of continuous monitoring on patient safety and quality of care, and effect on nurse-patient interaction. The interview guide is available on request. The interviews were conducted by two researchers with a biomedical and medical background and who were trained in interviewing.

Questionnaires

To determine psychological distress, all patients completed the short version of the State Trait Anxiety Inventory (STAI) [31,32] at baseline and on each day of the study period. On day 3, they completed the Pain Catastrophizing Scale (PCS), which provided a valid index about the extent to which people catastrophize [33]. STAI and PCS scores were compared between randomization groups as psychological distress can be a confounding factor. Furthermore, this allowed us to assess whether the devices affected psychological distress. In addition, nurses who took care of the participating patients and who were involved in, for example, attachment of the devices and changing batteries completed the System Usability Scale (SUS) [34], which is a reliable tool for assessing usability.

Analysis

Interviews

All interviews were audio-recorded and transcribed verbatim. Subsequently, two researchers (MW and TB) individually performed a thematic content analysis on all data (double coding) to determine facilitators, barriers, and positive and negative effects [35,36]. The researchers discussed the results until consensus was reached. The Donabedian framework for the quality of health care was used to present all positive and negative effects [37]. This framework distinguishes structure (context in which the care is delivered), process (all actions that make up health care), and outcome (all effects on patients' health). Facilitators and barriers were categorized according to an existing framework concerning determinants of adoption of mobile health [38,39]. New determinants regarding the use of VM and HP were added to the framework. Interviews were consecutively analyzed during the study, and saturation was assessed using histograms, in which all new factors per interview were presented. Quotes and striking issues were also documented. Once data saturation was reached, no further interviews were analyzed as it was expected that no new factors would be identified.

Questionnaires

STAI, PCS, and SUS scores were analyzed using SPSS package version 20.0 (SPSS Inc). STAI scores ranged from 6 to 24, and a higher score indicated more psychological distress. SUS scores ranged from 0 to 100, and a score above 68 was considered above average [34].

Descriptive statistics were presented as mean (SD). Statistical significance between patient groups regarding demographics and PCS was calculated using the analysis of variance (ANOVA) or Pearson chi-square test. ANOVA for repeated measures was used to assess differences in the STAI score between days and randomization groups. An independent samples *t* test was used to calculate the difference between HP and VM regarding SUS. STAI and SUS results were not correlated with the interview results. A *P* value less than .05 was considered significant.

Results

Demographics

A total of 165 patients were invited to participate: 89 patients from the surgical ward and 76 patients from the internal medicine ward. In each ward, 58 patients signed the informed consent form; 45 patients eventually participated in the study. Reasons for refusal were expectation of a large mental (n=37) or physical burden (n=10) and expected discharge within 24 hours (n=2). In the surgical ward, 13 patients were excluded because of rescheduling of the surgery (n=5), withdrawal of informed consent (n=4), early death (n=2), prolonged ICU stay (n=1), and a delirium (n=1). Reasons to exclude patients in the internal medicine ward were monitoring for less than 24 hours because of unexpected discharge (n=11) or physical burden by VM (n=2). No differences were found between randomization groups regarding age (*P*=.74) and gender (*P*=.55). Demographics are shown in Table 1. Relatives of 6 patients attended the interview. Six medical doctors (2 surgeons, 2 internists, and 2 intensivists), 3 physician assistants, and 20 nurses were interviewed.

Table 1. Patient demographics.

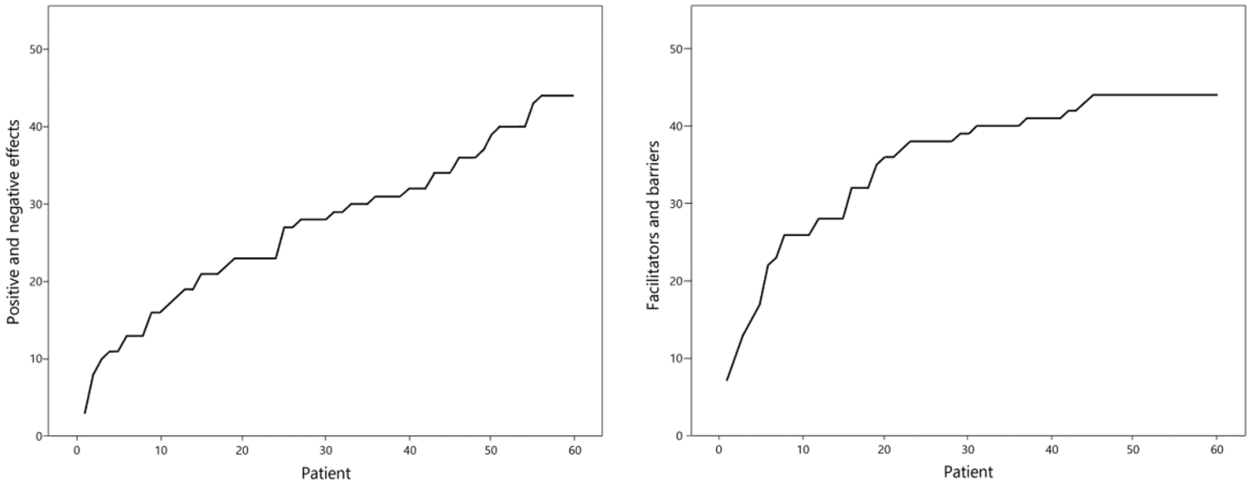
Demographics	ViSi Mobile (n=30)	HealthPatch (n=30)	Control group (n=30)
Gender, n (%)			
Male	18 (60)	22 (73)	20 (67)
Female	12 (40)	8 (27)	10 (33)
Age (years), median (range)	63 (26-76)	56 (27-88)	62 (34-77)
Measurement period (days) of participation in the study, median (range)	3 (1-4)	3 (1-5)	3 (2-3)
Reason for admission, n (%)			
Colorectal disease	8 (27)	8 (27)	5 (17)
Malignant	7 (23)	8 (27)	5 (17)
Benign	1 (5)	N/A[a]	N/A
Hepatobiliary disease	5 (17)	5 (17)	5 (17)
Malignant	5 (17)	2 (15)	5 (17)
Benign	N/A	3 (10)	N/A
Upper gastrointestinal disease			
Malignant	N/A	N/A	2 (7)
Neuroendocrine tumors			
Malignant	N/A	1 (3)	2 (7)
Herniation	1 (3)	1 (3)	1 (3)
Hematological diseases	N/A	1 (3)	2 (7)
Autoimmune diseases	4 (13)	2 (7)	N/A
Infectious disease	3 (10)	7 (23)	6 (20)
Other	9 (30)	5 (17)	7 (23)

[a] N/A: not applicable.

Interview Data

After analyzing 60 interviews (19 VM group, 21 HP group, and 20 control group) with patients, data saturation occurred, indicating that it was considered unlikely that new factors would be identified in additional interviews (Figure 1). We interviewed 29 health care professionals: 6 medical doctors, 3 physician assistants, and 20 nurses. After interviewing and analyzing professionals, we concluded that data saturation may not have been reached. Interviews with patients lasted for a median of 16 min (range 37 min). Generally, interviews with study participants from the control group lasted for a shorter duration as these people had no experience with the device. For professionals, interviews lasted for a median of 33 min (range 33 min). A total of 33 unique positive effects by patients and 56 positive effects by health care professionals and 14 negative effects by patients and 31 negative effects by health care professionals were identified. Patients reported 13 facilitators and 22 barriers, and health care professionals reported 13 facilitators and 36 barriers.

Figure 1. Saturation of positive and negative effects and facilitators and barriers. X-axis represents number of patients interviewed; Y-axis represents the accumulated number of new items mentioned by patients.



Positive Effects

In the structure, process, and outcome domains, 1, 23, and 23 positive effects were identified, respectively, (Multimedia Appendix 1) by patients, their relatives, and health care professionals. Moreover, 6 patients and 2 nurses mentioned alarms as a positive effect of continuous monitoring using wearable devices. A nurse stated:

> *We should all receive a mini-Ipad. It can show us patients' vital signs during our shift and will send us an alert in case the vital signs drop outside the normal ranges.* [NurseID4]

A total of 17 patients, 2 relatives, and 17 health care professionals expected to be able to detect clinical deterioration in an earlier phase using continuous monitoring. Moreover, 5 patients, 3 nurses, and 1 medical doctor mentioned that earlier detection can result in earlier interventions. Furthermore, 6 patients, 1 relative, and 5 nurses thought that the implementation of continuous monitoring can lead to less patient disturbances. In addition, 7 patients and 11 health care professionals thought that continuous monitoring can save time. We asked all nurses how to spend the saved time. A nurse mentioned:

> *Just talking to the patient. To have more time for the story of the patient.* [NurseID7]

Other positive effects regarding efficiency in health care were a reduced workload, shorter hospital stay, prevention of ICU admission, reduced costs, and lower amount of nursing staff required. A patient described:

> *You can stay shorter in the hospital and can go home with a wearable device. They can inspect your data in the hospital while you are at home. I would like that, it would feel more safe.* [PatientID40]

A total of 17 patients, 1 relative, and 9 health care professionals expected increased feelings of safety in patients in the general ward. In addition, patients' relatives and nurses mentioned feeling safer. A nurse explained:

> *Postoperative patients have been monitored continuously at the ICU. Some do feel unsafe after return at the general ward because of a lower number of vital sign measurements.* [NurseID1]

All nurses and most patients encouraged the implementation of wearable devices for continuous monitoring of patients. A nurse and a patient mentioned:

> *This is the future. We have to deal with it and the sooner we start working with those wearable devices, the more profit we will have.* [NurseID16]
>
> *The future...I think only 30% of the patients will be hospitalized by then. Patients will be monitored from home with this kind of smart devices.* [PatientID50]

Negative Effects

A total of 12 and 18 negative effects were identified in the process and outcome domains, respectively, by patients, their relatives, and health care professionals (Multimedia Appendix 2). Moreover, 1 patient and 5 health care professionals thought that continuous monitoring can generate an overload of information. An internist mentioned:

> *Sometimes you just do not want to know, making yourself crazy with too much data. Particularly when data does not influence your decision in patient's treatment.* [MedicaldoctorID5]

Particularly, nurses in the surgical ward were afraid that their ward would become like an ICU; 3 nurses and 1 medical doctor thought that this can lead to reluctance in transfer to the ICU. The alarm system was mentioned as a negative effect by 3 nurses and 3 medical doctors, leading to false-positive alarms, irrelevant alarms, and alarm fatigue. Moreover, 9 patients, 1 relative, and 5 nurses were afraid that interaction between patient and health care professionals would be reduced. A patient mentioned:

> *You need the confidence from the nurses, I would miss that. However, quantity time might become quality time.* [PatientID58]

In addition, 7 nurses and 1 medical doctor mentioned that continuous monitoring would cost more time, and 1 nurse, 1 physician assistant, and 1 medical doctor thought that it would increase workload. A nurse said:

> *Maybe it will increase work load. What if you receive an alarm every time a patient falls asleep and the oxygen saturation decreases a little bit?* [NurseID6]

A total of 12 patients, 2 relatives, and 2 health care professionals mentioned that patients can become worried by being able to see their own vital signs. A patient explained:

> *Some people are very anxious. Like my wife...like she already said: she would overreact. I would like to know my vital signs, but she would panic.* [PatientID54]

Facilitators

Eight facilitators were found in the domain Factors related to devices (Multimedia Appendix 3). In addition, 1 nurse and 1 medical doctor mentioned that by using continuous monitoring, health care professionals are able to see trends in vital signs. A surgeon stated:

> *Last night we saw a patient with an Early Warning Score of 3 and in the morning it suddenly was 13. Using continuous monitoring, we would have been able to see the Early Warning Score slowly increasing during the night.* [MedicaldoctorID6]

Moreover, 2 patients, 2 nurses, and 1 medical doctor mentioned the small size of the HP. Three patients said that they thought it was easy to view all vital signs on the VM wrist device or the mobile device of the HP. In addition, 2 patients and 1 nurse said that they think both the devices are reliable.

Three facilitators were found in the domain Individual factors. In addition, 2 patients, 2 nurses, and 2 medical doctors thought that continuous monitoring will lead to earlier detection of clinical deterioration, and 2 patients mentioned that they think that patient safety will be improved. In the Human environment domain, 8 facilitators were identified. Five patients mentioned that the devices were invisible under their clothes, and 7 patients said that they were not aware of the device. Moreover, 1 patient, 2 nurses, and 1 medical doctor mentioned fewer actions during vital sign measurements as a facilitator, such as putting on the upper arm cuff for BP measurements.

Barriers

In the domain Factors related to the devices, 22 barriers were identified (Multimedia Appendix 3). In addition, 2 patients, 3 nurses, and 1 medical doctor mentioned the VM battery change as a barrier. VM wrist device was thought to be too big or heavy by 5 patients, 3 nurses, and 1 medical doctor. Furthermore, VM cables and the patches and electrodes were also mentioned as barriers. A patient said:

> *Yesterday I felt very ill. I noticed that when you do not feel very well, every line, every device is just too much.* [PatientID40]

Three patients mentioned that devices are not able to measure patient experiences, such as pain. A patient described:

> *The devices are not able to register pain. When the nurse does not visit me, I cannot tell her I am having a headache. The device will not register that.* [PatientID55]

Moreover, 2 patients and 5 nurses said that it is a barrier that the HP is not able to measure all vital signs. Furthermore, it was also mentioned that VM and HP both are not able to measure core temperature.

Four barriers were identified in the Individual factors domain. One medical doctor mentioned the risk of overtreatment by identifying abnormalities in vital signs that cannot be ignored. Moreover, 1 medical doctor and 1 patient said that the VM wrist device is stigmatizing. In the domain Human environment, six barriers were identified. Three patients thought that it was a burden to carry the HP mobile device. One medical doctor feared that there will be too much attention for the vital signs and less attention for the individual patient. One nurse mentioned that patients were worried that the patches would come off. Four barriers were identified in the Organizational environment domain. Two medical doctors mentioned that nurses do not have adequate training to interpret continuous data. In addition, 4 nurses and 1 medical doctor thought that there would not be enough personnel to monitor all data:

> *At this moment it is not feasible to monitor all patients 24 hours a day and to anticipate adequately to clinical deterioration with the amount of nursing staff we have.* [NurseID6]

Questionnaires

Psychological Distress

No significant effect between the three randomization groups was found on STAI score (*P*=.33), and no significant within-subject effect was found in STAI score between days (*P*=.78; Table 2). Data of surgical and internal medicine patients were calculated separately; no significant effect between the randomization groups was found on STAI score (*P*=.86 and *P*=.17, respectively). No significant differences were found between the three randomization groups regarding PCS (*P*=.57; Table 2).

Table 2. State Trait Anxiety Inventory and Pain Catastrophizing Scale.

Group	STAI[a] baseline, mean (SD)	STAI day 1, mean (SD)	STAI day 2, mean (SD)	STAI day 3, mean (SD)	Pain Catastrophizing Scale, mean (SD)
ViSi Mobile	11.8 (2.7)	11.3 (2.9)	10.6 (2.6)	10.6 (3.0)	14.2 (11.2)
HealthPatch	11.4 (2.7)	11.2 (2.8)	11.5 (2.8)	11.2 (3.3)	15.7 (11.6)
Control	11.0 (3.1)	11.1 (3.1)	11.2 (3.3)	11.7 (3.5)	17.4 (10.9)

[a] STAI: State Trait Anxiety Inventory.

Usability

The SUS was filled in by 6 nurses (3 internal medicine nurses and 3 surgical nurses), 1 for each device. Both devices scored above average, indicating good usability. No significant difference was found between VM and HP (mean 77.9, SD 18.5 and mean 82.5, SD 18.6, respectively; *P*=.68).

Discussion

Principal Findings

In this study, we used two wearable devices for continuous monitoring of vital signs in non-ICU patients with a wide spectrum of clinical conditions in two different wards. Our study resulted in a broad overview of experiences and expectations with the devices of both patients and health care professionals. We showed that continuous monitoring in the ward was not only well received by most patients and their relatives but also by their health care professionals. We also identified relevant barriers of continuous monitoring with wearable devices and that using wearable devices did not affect stress levels. Both patients and health care professionals expected that continuous monitoring of vital signs would lead to an earlier identification of clinical deterioration and to an improvement of quality, safety, and efficiency in health care. We also identified relevant barriers of continuous monitoring with wearable devices.

Our semistructured interviews revealed a primarily positive attitude toward continuous monitoring. A recent study by Abelson et al [40] also confirmed that surgical patients have a positive attitude toward wearable devices and mobile apps and that they are willing to use them. Earlier detection of clinical deterioration was frequently mentioned by patients and health care professionals corresponding with the findings from a recent review by Cardona-Morrell et al [10]. They showed that continuous monitoring of vital signs in the general ward leads to an earlier detection of clinical deterioration [10]. Respondents mentioned that continuous monitoring could lead to saved time and reduced workload for nurses, which is also found in other studies [10]. All nurses mentioned that they would use this time for the patient, such as mobilization, washing or showering patients, providing information, and providing a listening ear for the patient. This might solve the problem for less nurse-patient interaction, which was frequently mentioned by patients. Future research should shed light on changes in nurses' workload after the implementation of continuous monitoring. A frequently reported barrier was the wrist device and cables of VM. Particularly, surgical patients mentioned that the VM cables were a burden in combination with other lines, such as abdominal drains and urinary catheters. However, patients did not feel restricted during daily activities. This is important as early appropriate mobilization improves recovery and reduces the risk of complications [41,42]. STAI and PCS scores revealed no differences in psychological distress between patients in the intervention and control groups, indicating that neither the VM nor the HP caused additional stress or reduced stress. According to SUS scores, the larger VM wrist device and cables did not influence the overall usability of the VM in comparison with the much smaller HP. It is expected that future devices will become even smaller while being able to wirelessly monitor an increasing number of vital signs continuously. The amount of data that will become available by continuous monitoring was mentioned as a negative effect by health care professionals, as it was expected that they would never be able to review all data. Big data analytics are available for effective storage and processing of large amounts of data [43,44]. Alarms can alert the nurse when patient's vital signs drop out of normal ranges, resulting in a high number of false-positive or irrelevant alarms or even alarm fatigue [27,45]. Machine learning algorithms can prevent unnecessary diagnostic procedures and overtreatment because of a reduced number of irrelevant and false-positive alarms [46-48].

Other Research

Few studies regarding continuous monitoring in the general ward exist. Brown et al [49] compared continuous monitoring using the EarlySense system with intermittent monitoring in a medical-surgical ward. This system includes a flat sensor that is placed under the patient's bed and monitors HR and respiratory rate continuously. They found a reduced number of days in the ICU and shorter overall hospital stay because of earlier interventions in patients who were monitored continuously. However, the system is not able to monitor other vital signs such as BP, SpO_2, and temperature when patients are out of bed. Other researchers used patches to monitor several vital signs such as the HP in this study, with promising results [50]. Using HP and VM, patients are able to mobilize throughout the hospital while being monitored continuously for relevant vital signs. VM measures almost all vital signs, which are required to calculate the Modified Early Warning Score and judge the clinical situation of the patient. More recent work found that patient monitoring systems should be tailored to users' needs [51].

Strengths and Limitations

An important strength of this study is that we were able to monitor patients admitted for various reasons for a longer period in a clinical setting in two different departments. We conducted a large number of semistructured interviews with both patients and health care professionals and were able to reach data saturation in patients about all predefined categories. This resulted in a comprehensive overview of the positive and negative effects of continuous monitoring and facilitators and barriers regarding VM and HP. The control group allowed us to collect current experiences from patients that were not yet influenced by using wearable devices. Regarding interviews with health care providers, data may not have saturated, although this is difficult to assess [28]. Selection bias could have occurred as not all invited patients participated in the study, particularly in the surgical ward. However, we randomized all patients to VM, HP, or a control group, and no significant differences were found between randomization groups, for example, gender, which minimized bias. Patients who refused to participate reported that they feared the mental or physical burden, particularly severely ill patients or patients with psychological distress. No differences were found in stress experienced between different randomization groups. Although the STAI questionnaire is validated for measuring psychological distress, many other stressful factors may have an impact on patients and may potentially influence the outcomes (stress before surgical procedures or complications during hospitalization).

Future Perspectives

Implementation of wearable devices for continuous monitoring is expected to influence health care in several ways. Patient safety can be improved as trained and experienced personnel can be warned during an earlier phase of deterioration and perform early interventions. This can prevent unnecessary ICU admissions and shorten hospitality stay. Nurses will have to be trained in using wearable devices and continuous vital sign data in the general ward. It is expected that nurses will have more time for other needs of a patient during admission. Data transmission via Wi-Fi between the device and the electronic health record should be safe and accurate. Potential alarms in vital signs can be processed using predictive analytics and machine learning techniques to prevent false-positive alarming. Furthermore, patients can benefit from continuation of monitoring using the same or comparable wearable devices. Vital signs data collected at home can be shared with trained nurses or physicians. With continuous monitoring, patients can be more actively involved in their own treatment. To stimulate this, the facilitators and barriers reported in this study are of great value when planning to implement wearable devices in the general ward.

Conclusions

According to patients and health care professionals, VM and HP have potential for continuous monitoring of vital signs in the general ward, and almost all of them encouraged the idea of monitoring vital signs continuously in the general ward. The comprehensive overview of barriers and facilitators of using wireless devices should be taken into consideration when choosing the device for implementing continuous monitoring. Continuous monitoring may facilitate the use of predictive analytics for clinical deterioration and early interventions. Further studies should explore the effect of continuous monitoring on clinical outcomes of patients in the general ward.

Acknowledgments

The project was funded by the Radboud University Medical Center.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Positive effects.
[DOCX File , 21 KB - jmir_v22i6e15471_app1.docx]

Multimedia Appendix 2
Negative effects.
[DOCX File , 18 KB - jmir_v22i6e15471_app2.docx]

Multimedia Appendix 3
Facilitators and barriers.
[DOCX File , 22 KB - jmir_v22i6e15471_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V. 1.6.1).
[PDF File (Adobe PDF File), 391 KB - jmir_v22i6e15471_app4.pdf]

References

1. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
2. Appelboom G, Camacho E, Abraham ME, Bruce SS, Dumont EL, Zacharia BE, et al. Smart wearable body sensors for patient self-assessment and monitoring. Arch Public Health 2014;72(1):28 [FREE Full text] [doi: 10.1186/2049-3258-72-28] [Medline: 25232478]
3. Iqbal MH, Aydin A, Brunckhorst O, Dasgupta P, Ahmed K. A review of wearable technology in medicine. J R Soc Med 2016 Oct;109(10):372-380 [FREE Full text] [doi: 10.1177/0141076816663560] [Medline: 27729595]
4. Chan M, Estève D, Fourniols J, Escriba C, Campo E. Smart wearable systems: current status and future challenges. Artif Intell Med 2012 Nov;56(3):137-156. [doi: 10.1016/j.artmed.2012.09.003] [Medline: 23122689]
5. Hentschel MA, Haaksma ML, van de Belt TH. Wearable technology for the elderly: underutilized solutions. Eur Geriatr Med 2016;7(5):399-401. [doi: 10.1016/j.eurger.2016.07.008]

6. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. JMIR Mhealth Uhealth 2017 Jul 5;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
7. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. Conf Proc IEEE Eng Med Biol Soc 2013;2013:4058-4061. [doi: 10.1109/EMBC.2013.6610436] [Medline: 24110623]
8. Sahandi R, Noroozi S, Roushan G, Heaslip V, Liu Y. Wireless technology in the evolution of patient monitoring on general hospital wards. J Med Eng Technol 2010 Jan;34(1):51-63. [doi: 10.3109/03091900903336902] [Medline: 19929237]
9. Goldberg EM, Levy PD. New approaches to evaluating and monitoring blood pressure. Curr Hypertens Rep 2016 Jun;18(6):49. [doi: 10.1007/s11906-016-0650-9] [Medline: 27137524]
10. Cardona-Morrell M, Prgomet M, Turner RM, Nicholson M, Hillman K. Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. Int J Clin Pract 2016 Oct;70(10):806-824. [doi: 10.1111/ijcp.12846] [Medline: 27582503]
11. Majumder S, Mondal T, Deen MJ. Wearable sensors for remote health monitoring. Sensors (Basel) 2017 Jan 12;17(1):pii: E130 [FREE Full text] [doi: 10.3390/s17010130] [Medline: 28085085]
12. Beckett DJ, Inglis M, Oswald S, Thomson E, Harley W, Wilson J, et al. Reducing cardiac arrests in the acute admissions unit: a quality improvement journey. BMJ Qual Saf 2013 Dec;22(12):1025-1031 [FREE Full text] [doi: 10.1136/bmjqs-2012-001404] [Medline: 23828879]
13. Haller G, Myles PS, Wolfe R, Weeks AM, Stoelwinder J, McNeil J. Validity of unplanned admission to an intensive care unit as a measure of patient safety in surgical patients. Anesthesiology 2005 Dec;103(6):1121-1129. [doi: 10.1097/00000542-200512000-00004] [Medline: 16306722]
14. Frost SA, Alexandrou E, Bogdanovski T, Salamonson Y, Parr MJ, Hillman KM. Unplanned admission to intensive care after emergency hospitalisation: risk factors and development of a nomogram for individualising risk. Resuscitation 2009 Feb;80(2):224-230. [doi: 10.1016/j.resuscitation.2008.10.030] [Medline: 19084319]
15. Calzavacca P, Licari E, Tee A, Egi M, Downey A, Quach J, et al. The impact of Rapid Response System on delayed emergency team activation patient characteristics and outcomes--a follow-up study. Resuscitation 2010 Jan;81(1):31-35. [doi: 10.1016/j.resuscitation.2009.09.026] [Medline: 19854557]
16. Kaboli PJ, Rosenthal GE. Delays in transfer to the ICU: a preventable adverse advent? J Gen Intern Med 2003 Feb;18(2):155-156 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.21217.x] [Medline: 12542592]
17. Beckett D, Gordon C, Paterson R, Chalkley S, Macleod D, Bell D. Assessment of clinical risk in the out of hours hospital prior to the introduction of Hospital at Night. Acute Med 2009;8(1):33-38. [Medline: 21607208]
18. Mapp ID, Davis LL, Krowchuk H. Prevention of unplanned intensive care unit admissions and hospital mortality by early warning systems. Dimens Crit Care Nurs 2013;32(6):300-309. [doi: 10.1097/DCC.0000000000000004] [Medline: 24100432]
19. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. Chest 2013 Jun;143(6):1758-1765 [FREE Full text] [doi: 10.1378/chest.12-1605] [Medline: 23732586]
20. Cretikos M, Chen J, Hillman K, Bellomo R, Finfer S, Flabouris A, MERIT study investigators. The objective medical emergency team activation criteria: a case-control study. Resuscitation 2007 Apr;73(1):62-72. [doi: 10.1016/j.resuscitation.2006.08.020] [Medline: 17241732]
21. Kyriacos U, Jelsma J, Jordan S. Record review to explore the adequacy of post-operative vital signs monitoring using a local modified early warning score (mews) chart to evaluate outcomes. PLoS One 2014;9(1):e87320 [FREE Full text] [doi: 10.1371/journal.pone.0087320] [Medline: 24498075]
22. Ludikhuize J, Borgert M, Binnekade J, Subbe C, Dongelmans D, Goossens A. Standardized measurement of the Modified Early Warning Score results in enhanced implementation of a Rapid Response System: a quasi-experimental study. Resuscitation 2014 May;85(5):676-682. [doi: 10.1016/j.resuscitation.2014.02.009] [Medline: 24561029]
23. Boatin AA, Wylie BJ, Goldfarb I, Azevedo R, Pittel E, Ng C, et al. Wireless vital sign monitoring in pregnant women: a functionality and acceptability study. Telemed J E Health 2016 Jul;22(7):564-571. [doi: 10.1089/tmj.2015.0173] [Medline: 27172448]
24. Zubiete ED, Luque LF, Rodríguez AV, González IG. Review of wireless sensors networks in health applications. Conf Proc IEEE Eng Med Biol Soc 2011;2011:1789-1793. [doi: 10.1109/IEMBS.2011.6090510] [Medline: 22254675]
25. Ohashi K, Kurihara Y, Watanabe K, Ohno-Machado L, Tanaka H. Feasibility evaluation of Smart Stretcher to improve patient safety during transfers. Methods Inf Med 2011;50(3):253-264. [doi: 10.3414/ME0616] [Medline: 21057715]
26. Welch J, Moon J, McCombie S. Early detection of the deteriorating patient: the case for a multi-parameter patient-worn monitor. Biomed Instrum Technol 2012;Suppl:57-64. [doi: 10.2345/0899-8205-46.s2.57] [Medline: 23039778]
27. Clifton L, Clifton DA, Pimentel MA, Watkinson PJ, Tarassenko L. Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. IEEE J Biomed Health Inform 2014 May;18(3):722-730. [doi: 10.1109/JBHI.2013.2293059] [Medline: 24808218]
28. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
29. Selvaraj N. Long-Term Remote Monitoring of Vital Signs Using a Wireless Patch Sensor. In: Proceedings of the 2014 IEEE Healthcare Innovation Conference. 2014 Presented at: HIC'14; October 8-10, 2014; Seattle, WA, USA.

30. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004 Mar 15;180(S6):S57-S60. [Medline: 15012583]
31. Speilberger C, Gorsuch R, Lushene R. STAI manual. Palo Alto, CA: Consulting Psychologist Press; 1970.
32. van der Bij AK, de Weerd S, Cikot RJ, Steegers EA, Braspenning JC. Validation of the dutch short form of the state scale of the Spielberger State-Trait Anxiety Inventory: considerations for usage in screening outcomes. Community Genet 2003;6(2):84-87. [doi: 10.1159/000073003] [Medline: 14560068]
33. Sullivan MJ, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
34. Brooke J. SUS - A quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London: Taylor & Francis; 1996:189-194.
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
37. Donabedian A. The quality of care. How can it be assessed? J Am Med Assoc 1988;260(12):1743-1748. [doi: 10.1001/jama.260.12.1743] [Medline: 3045356]
38. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
39. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
40. Abelson JS, Symer M, Peters A, Charlson M, Yeo H. Mobile health apps and recovery after surgery: What are patients willing to do? Am J Surg 2017 Oct;214(4):616-622. [doi: 10.1016/j.amjsurg.2017.06.009] [Medline: 28666581]
41. Markey DW, Brown RJ. An interdisciplinary approach to addressing patient activity and mobility in the medical-surgical patient. J Nurs Care Qual 2002 Jul;16(4):1-12. [doi: 10.1097/00001786-200207000-00002] [Medline: 12125898]
42. Pashikanti L, Von Ah D. Impact of early mobilization protocol on the medical-surgical inpatient population: an integrated review of literature. Clin Nurse Spec 2012;26(2):87-94. [doi: 10.1097/NUR.0b013e31824590e6] [Medline: 22336934]
43. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
44. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
45. Welch J, Kanter B, Skora B, McCombie S, Henry I, McCombie D, et al. Multi-parameter vital sign database to assist in alarm optimization for general care units. J Clin Monit Comput 2016 Dec;30(6):895-900 [FREE Full text] [doi: 10.1007/s10877-015-9790-8] [Medline: 26439830]
46. Eerikäinen LM, Vanschoren J, Rooijakkers MJ, Vullings R, Aarts RM. Reduction of false arrhythmia alarms using signal selection and machine learning. Physiol Meas 2016 Aug;37(8):1204-1216. [doi: 10.1088/0967-3334/37/8/1204] [Medline: 27454128]
47. Antink CH, Leonhardt S, Walter M. Reducing false alarms in the ICU by quantifying self-similarity of multimodal biosignals. Physiol Meas 2016 Aug;37(8):1233-1252. [doi: 10.1088/0967-3334/37/8/1233] [Medline: 27454256]
48. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
49. Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. Am J Med 2014 Mar;127(3):226-232. [doi: 10.1016/j.amjmed.2013.12.004] [Medline: 24342543]
50. Breteler MJ, Huizinga E, van Loon K, Leenen LP, Dohmen DA, Kalkman CJ, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. BMJ Open 2018 Feb 27;8(2):e020162 [FREE Full text] [doi: 10.1136/bmjopen-2017-020162] [Medline: 29487076]
51. Poncette AS, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, et al. Clinical requirements of future patient monitoring in the intensive care unit: qualitative study. JMIR Med Inform 2019 Apr 30;7(2):e13064 [FREE Full text] [doi: 10.2196/13064] [Medline: 31038467]

Abbreviations

ANOVA: analysis of variance
BP: blood pressure
ECG: electrocardiography
HP: HealthPatch
HR: heart rate

ICU: intensive care unit
PCS: Pain Catastrophizing Scale
SpO_2: oxygen saturation
STAI: State Trait Anxiety Inventory
SUS: System Usability Scale
VM: ViSi Mobile

Edited by G Eysenbach; submitted 12.07.19; peer-reviewed by A Boatin, AS Poncette, JW Kim, C Subbe; comments to author 21.10.19; revised version received 09.01.20; accepted 24.01.20; published 10.06.20.

Please cite as:
Weenk M, Bredie SJ, Koeneman M, Hesselink G, van Goor H, van de Belt TH
Continuous Monitoring of Vital Signs in the General Ward Using Wearable Devices: Randomized Controlled Trial
J Med Internet Res 2020;22(6):e15471
URL: https://www.jmir.org/2020/6/e15471
doi:10.2196/15471
PMID:32519972

Original Paper

Internet Influence of Assisted Reproduction Technology Centers in China: Qualitative Study Based on WeChat Official Accounts

Fang Shao[1], PhD; Zhiqiang He[1], MPhil; Zheng Zhu[2], MBBS; Xiang Wang[1], BSc; Jianping Zhang[3], BSc; Jinhua Shan[3], BSc; Jiajia Pan[4], BSc; Hui Wang[5,6], PhD

[1]Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China
[2]School of Pediatrics, Nanjing Medical University, Nanjing, China
[3]School of Nursing, Nanjing Medical University, Nanjing, China
[4]School of Public Health, Nanjing Medical University, Nanjing, China
[5]Department of Histology and Embryology, School of Basic Medical Sciences, Nanjing Medical University, Nanjing, China
[6]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China

Corresponding Author:
Hui Wang, PhD
Department of Histology and Embryology, School of Basic Medical Sciences
Nanjing Medical University
101 Longmian Rd
Nanjing, 211166
China
Phone: 86 025 86869380
Email: anniewang@njmu.edu.cn

Abstract

Background: The prevalence of infertility in China is high, but the advent of assisted reproduction technology (ART) has greatly eased this situation. Social media, such as WeChat official accounts, have become the preferred tool for ART centers to communicate with patients, but their attention and operational status differ, and the Internet influence of WeChat official accounts is insufficient. In addition, questions about whether Internet influence is consistent with academic influence and whether the Internet can influence patients' choice of medical treatment to a certain extent have not been explored.

Objective: This study aimed to examine the operational status and Internet influence of WeChat official accounts for ART centers and to explore the degree of Internet influence on patients' choices of medical treatment.

Methods: We collected information from the WeChat official accounts for ART centers approved by the National Health Commission of the People's Republic of China and used the technique for order of preference by similarity to ideal solution to build an Internet influence model of the ART centers and obtained a Ranking of Internet Influence on Reproductive Centers (RIIRC) for each center.

Results: We found there were 451 ART centers throughout the country by the end of 2016 and 498 by the end of 2018. The number of medical institutions is quite large, but their distribution is uneven, and their level of medical technical ability is very different. Analysis of the text data of posts of WeChat official accounts showed the ART centers have insufficient awareness of network exposure and publicity, and the RIIRC of some medical institutions was inconsistent with their medical level and academic status.

Conclusions: ART institutions have varying degrees of emphasis and use of WeChat official accounts in China. They fail to realize that the Internet influence of WeChat may bring them potential patient resources and that Internet influence may affect the future market structure of ART and may also potentially affect academic rankings.

(J Med Internet Res 2020;22(6):e17997) doi:10.2196/17997

KEYWORDS

ART center; WeChat official account; Internet influence

Introduction

Severe Infertility and the Increased Requirements of Assisted Reproduction Technology Service

The prevalence of infertility in China, which has increased from 6.7% in 1988 to 15% to 20% now [1,2], has been influenced by a number of factors, including marital status, educational level, delayed age of child-bearing, the two-child policy, repeatedly induced abortions, poor lifestyle, and environmental pollution. Infertility has brought a heavy burden on countless families and has had a number of important effects on both personal and public health [3-6], including damage to social reputation [7], increased psychological pressure [8], long-term infertility [9,10], and constraints on economic development [11].Infertility is not only a physical condition but also a complicated sociological problem in China. However, it is possible to achieve pregnancy through assisted reproductive technology (ART) in cases that cannot be successfully addressed by conventional drugs or surgery.

The National Health Commission of the People’s Republic of China (NHCPRC) approved the implementation of human ART licenses in medical institutions more than 30 years ago. However, it was not until 2007 that the State Council delegated approval authority to the provincial health commissions. Before then, there were only 95 ART centers in China [12], and the number of these centers was initially limited, and their growth was slow because of the technical barriers to ART. Since the decentralization of approval authority and the development of the second and third generations of ART, these centers grew rapidly, and by the end of 2018, there were 498 such centers in China [13]. However, this rapid growth has led to uneven development in terms of the geographical distribution and the level of medical technology.

Competition in Assisted Reproduction Technology Market Approaches into the Qualitative Stage

In 2015, the NHCPRC proposed there should be *one institution for every 3 million people* to plan the development of ART centers [14]. The total number of ART centers in China under this policy was set at about 550, and the number of ART centers is approaching this limit, which is expected to result in a new stage of qualitative growth rather than quantitative growth. As it is extremely unlikely that there will be new public medical institutions, private capital investment is making competition in the ART market increasingly fierce. Many private investors have entered into the 100 billion market, such as Chengdu Xi’nan Gynecology Hospital and China IVF Medical Group Medical Group. Therefore, the focus of the ART centers in this next stage should be not only to upgrade their medical technology but to expand their public presence to attract the attention of potential patients.

Ignorance of Internet Images Influencing Patients’ Choice by Assisted Reproduction Technology Centers

Recent statistics indicate that the number of mobile internet users in China has reached 847 million, of which 48.3% are between 20 and 40 years of age [15]—the so-called generation of “Cyberspace natives.” They are accustomed to searching for information on the internet and forming their own opinions, which guides their decision in selecting medical institutions, but they may not pay much attention to the academic standing of medical institutions. This age group is also the main population that is interested in pregnancy preparation, and many of them are currently undergoing ART.

In contrast, most hospitals, including their ART centers, have ignored the impact of the internet, focusing instead on feedback from patients. In the highly competitive medical market, the technology and equipment of ART centers are increasingly easy to imitate and replace, and it is difficult for their hospitals to create their own unique identity. As the costs of infertility diagnosis and treatment are not covered by essential national medicare and the role of Cyberspace natives, internet influence on ART centers cannot be ignored. Is the influence of the internet on the ART center consistent with the influence of academic standing? What is the gap between them? Whether the internet influences patients' choices, and whether it can attract the attention of the capital market and alter the nature of hospitals in the future have not been examined.

The Aim of This Study

This study investigated the operation and popularization of the WeChat internet connection by ART centers. We used the technique for order of preference by similarity to ideal solution (TOPSIS) to construct an internet influence measure of ART centers and obtain a Ranking of Internet Influence on Reproductive Centers (RIIRC), which provides guidance for encouraging patients to seek reproductive help at ART centers by enhancing public awareness of them and their public image.

Methods

Sample Selection

The initial study sample consisted of 451 ART centers that had been approved by the NHCPRC by the end of 2016. A total of 498 medical institutions were approved by the NHCPRC as of December 31, 2018 [13]. The 47 new medical institutions approved between December 31, 2016, and December 31, 2018, were not included in the study because of their short history.

We began by browsing the official website of each ART center, including its main page and subpages, searching for the WeChat icon, QR codes, and opening announcements. If no WeChat official account was found, we proceeded to the next step, which was to search for sites that contained words or phrases such as “Weixin” or “WeChat.” If the website of the ART center did not have a “search” function, we used Google or Baidu to search within the medical institution’s website.

We also searched directly in the WeChat App using the keywords “name of medical institution,” “reproductive medicine,” “ART center,” or “reproductive medicine department.” Finally, we consulted with the person in charge by emails, messages on the website of ART center, etc. If we confirmed that WeChat account have been applied, they will not be consulted.

These procedures revealed that 200 of the 451 ART centers had opened WeChat official accounts, as of December 31, 2018.

Afterwards, Webpage parsing techniques and other technical means were used to obtain basic information for the 200 centers.

The WeChat accounts of the 200 ART centers were screened again, and unqualified accounts were excluded. We examined the text data of the official account posts to see if they met the first exclusion criterion: that is, having multiple values (five or more values) on the 19 indicators described in the section on Index of Internet Influence. A total of 41 accounts that had multiple missing values were excluded. The second criterion was that there was no record value; that is, the account had not been continuously maintained and updated. We found that nine accounts had not pushed tweets or only had a few tweets in the past year, so they were excluded. This left a total of 150 official accounts included in the analysis. All NULL values in the text of tweets of the 150 accounts were replaced with zero for the convenience of subsequent processing.

Collection of Basic Data

The basic data of the 150 centers were collected manually to evaluate the influence of the internet through the WeChat accounts. We hired 4 assistant researchers to inspect and collect the data after they received professional training. The account data from each center were randomly divided into four equal parts, and the 4 assistant researchers randomly selected one of them to inspect and collect the data. After one round of collection, they exchanged the list for a second round, and finally inspected the two rounds of data. If a discrepancy was found, it was discussed and consensus was reached.

Basic information included an assessment index for the access to technology, an official account menu bar score, geographical location, date of formation on the department, date of approval by the NHCPRC, date of creating the WeChat official account, and whether there was an official website and official hospital certification (service account and subscription account).

The assessment index for access to technology refers to the type of ART approved by the NHCPRC (Table 1).The official account menu bar score refers to the functions contained in the menu bar of the official account, which generally included a department introduction, appointment registration, navigation, medical examination reports, popular scientific articles, etc. One point was assigned for each function and the sum of the points was the score.

Table 1. Types and scores of assisted reproduction technology in medical institutions.

Access to technology	Score
Artificial insemination with husband (AIH)	0.5
Artificial insemination by donor (AID)	0.5
In vitro fertilization and embryo transfer (IVF-VT)	2
Intracytoplasmic sperm injection (ICSI)	3
Preimplantation genetic diagnosis (PGD)	4

Acquisition, Conversion, and Cleaning of Data

Data mining of WeChat accounts generally requires using appropriate methods and tools, according to the characteristics of the dataset. This study mainly adopted the Webpage parsing technique, using Python 3.7 as the development language, PyChram as the development platform, and MongoDB for data storage. The data had to be further processed and structured for later analysis and mining, mainly involving data deduplication, integration, and transformation.

The data from the eligible centers were from November 1, 2018 to October 31, 2019.

Index of Internet Influence

In light of the existing research literature and the purpose of this study [16-18], we developed an index to measure the internet influence for ART centers through their WeChat accounts. The 19 indices shown in Table 2 were selected for use in the study.

It was necessary to clarify the concepts and definitions of indices selected in this study. WeChat service account and subscription account can setup to eight tweets at a time. The heading of the main title was called the “headline,” and the seven subtitles were collectively referred to as “subheadlines.” “Posting frequency” (x_1, number of posts per month) equals to the number of articles WeChat account posts in the past year divided by 12. “Number of articles of each post” (x_5, number of articles per post) equals to the total number of articles WeChat accounts posts divided by the number of posts in the past year. “Average number of views of the headlines” (x_6, number of views per article) equals to the total number of views of headline articles divided by the number of headlines in the past year. “Average number of likes of headlines” (x_7, number of likes per article) equals to the total number of likes of the headline articles divided by the number of headlines in the past year. The number of views and likes of posts of the other indicators used a calculation method similar to that of the “Average number of views of the headlines” and “Average number of likes of headlines.”

Table 2. The index of internet influence for assisted reproductive technology centers.

Category and index	Symbol
Basic index	
Posting Frequency	x_1
Number of articles of each post	x_5
Menu bar score	x_{18}
Headline index	
Average number of views of headlines	x_6
Average number of likes of headlines	x_7
Average number of views of original headlines	x_8
Average number of likes of original headlines	x_9
Average number of views of unoriginal headlines	x_{10}
Average number of likes of unoriginal headlines	x_{11}
Academic index	
Assessment Index for Access Technology	x_{19}
Articles type	
Number of headlines	x_2
Number of subheadlines	x_3
Number with original content	x_4
Subheadline index	
Average number of views of subheadlines	x_{12}
Average number of likes of subheadlines	x_{13}
Original content	
Average number of views of original content	x_{14}
Average number of likes of original content	x_{15}
Average number of views of unoriginal content	x_{16}
Average number of likes of unoriginal content	x_{17}

Construction of the Internet Influence Measure

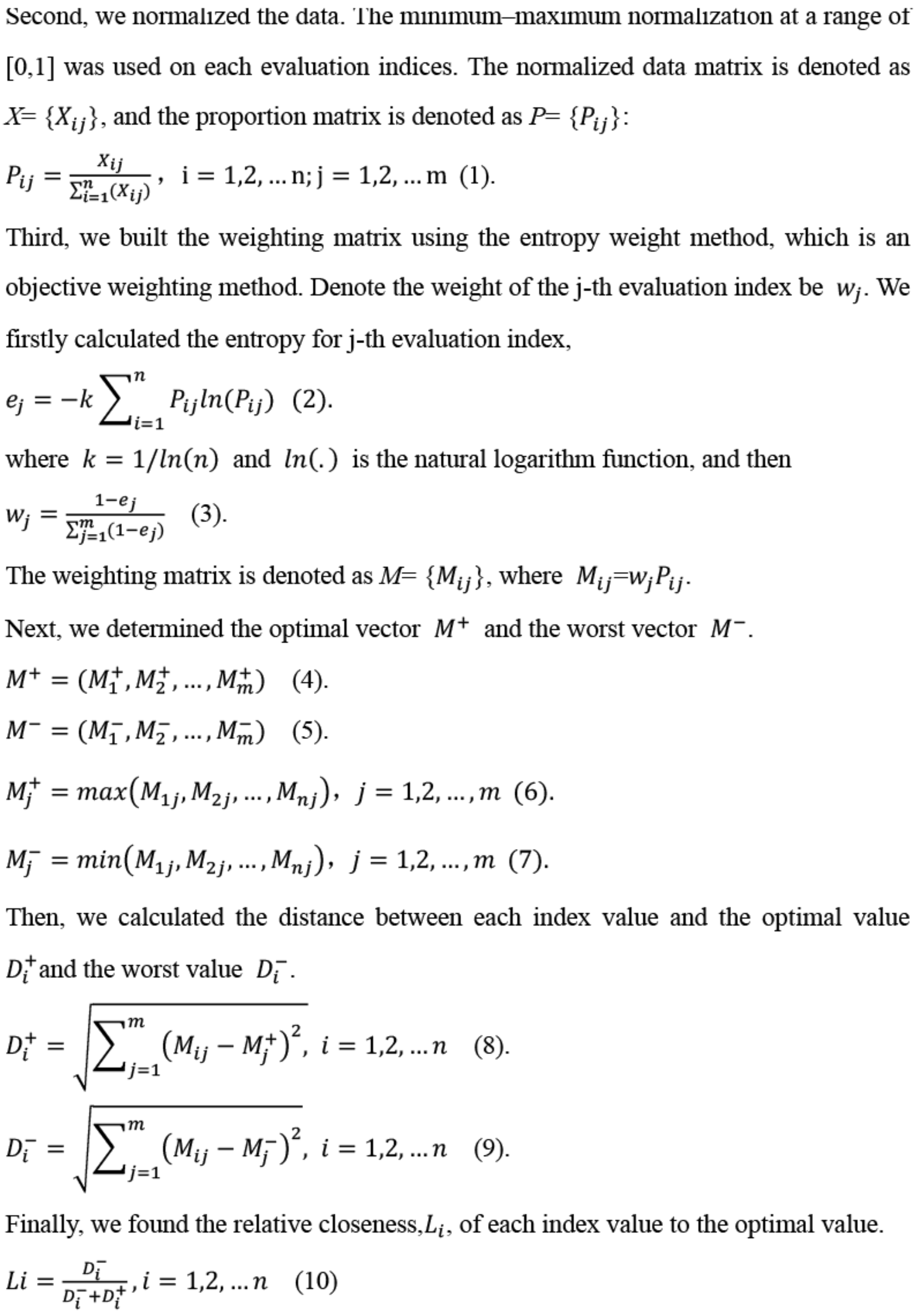
The construction of the measure of internet influence and the output of the figures were based on R Programming Language (The R Foundation Conference Committee, version 3.6.1, 2019-07-05).

The determination of the weight of the index and measurement model were based on TOPSIS.

First, we created an index matrix; there were n evaluation objects, and each evaluation object had m evaluation indices. It is usually necessary to convert low-quality indices with lower values and better evaluation results to high-quality indices to make the objectives have the same trend. In this study, however, all of the indices were high-quality indices, so conversion did not have to be performed.

The proportion matrix P_{ij}, weighting matrix M_{ij}, and relative closeness Li were computed by equations (1), (4), (5), (6), (7), and (10) as shown in Figure 1.

Figure 1. Mathematical equations.



Second, we normalized the data. The minimum–maximum normalization at a range of [0,1] was used on each evaluation indices. The normalized data matrix is denoted as $X = \{X_{ij}\}$, and the proportion matrix is denoted as $P = \{P_{ij}\}$:

$$P_{ij} = \frac{X_{ij}}{\sum_{i=1}^{n}(X_{ij})}, \quad \mathrm{i} = 1,2, \ldots \mathrm{n}; \mathrm{j} = 1,2, \ldots \mathrm{m} \quad (1).$$

Third, we built the weighting matrix using the entropy weight method, which is an objective weighting method. Denote the weight of the j-th evaluation index be w_j. We firstly calculated the entropy for j-th evaluation index,

$$e_j = -k \sum_{i=1}^{n} P_{ij} ln(P_{ij}) \quad (2).$$

where $k = 1/ln(n)$ and $ln(.)$ is the natural logarithm function, and then

$$w_j = \frac{1-e_j}{\sum_{j=1}^{m}(1-e_j)} \quad (3).$$

The weighting matrix is denoted as $M = \{M_{ij}\}$, where $M_{ij} = w_j P_{ij}$.

Next, we determined the optimal vector M^+ and the worst vector M^-.

$$M^+ = (M_1^+, M_2^+, \ldots, M_m^+) \quad (4).$$

$$M^- = (M_1^-, M_2^-, \ldots, M_m^-) \quad (5).$$

$$M_j^+ = max(M_{1j}, M_{2j}, \ldots, M_{nj}), \quad j = 1,2, \ldots, m \quad (6).$$

$$M_j^- = min(M_{1j}, M_{2j}, \ldots, M_{nj}), \quad j = 1,2, \ldots, m \quad (7).$$

Then, we calculated the distance between each index value and the optimal value D_i^+ and the worst value D_i^-.

$$D_i^+ = \sqrt{\sum_{j=1}^{m} (M_{ij} - M_j^+)^2}, \quad i = 1,2, \ldots n \quad (8).$$

$$D_i^- = \sqrt{\sum_{j=1}^{m} (M_{ij} - M_j^-)^2}, \quad i = 1,2, \ldots n \quad (9).$$

Finally, we found the relative closeness, L_i, of each index value to the optimal value.

$$Li = \frac{D_i^-}{D_i^- + D_i^+}, i = 1,2, \ldots n \quad (10)$$

Results

Distribution and Technical Ability of the Centers Is Uneven

The 451 ART centers were located in 31 provincial administrative regions of China, including 22 provinces, five autonomous regions, and four municipalities directly under the central government. However, they were mainly in the southeast coast, the middle and lower reaches of the Yangtze River, and the Bohai Rim (Figure 2).

Nearly three-quarters (327/451, 72.5%) of the ART centers in Chinese performed in vitro fertilization-embryo transfer and intracytoplasmic sperm injection, whereas only 8.9% (40/451) could perform preimplantation genetic diagnosis (Table 3).

The Third Hospital of Beijing Medical University (now known as Peking University Third Hospital) and the Second Clinical Medical College of Hunan Medical University (now known as Xiangya Hospital Central South University) were the earliest medical institutions to conduct ART in China. The number of ART centers increased dramatically from 2007 to 2016, during which time a total of 403 ART centers were approved (Table 4).

Figure 2. Distribution of 451 assisted reproduction technology center across Mainland China.

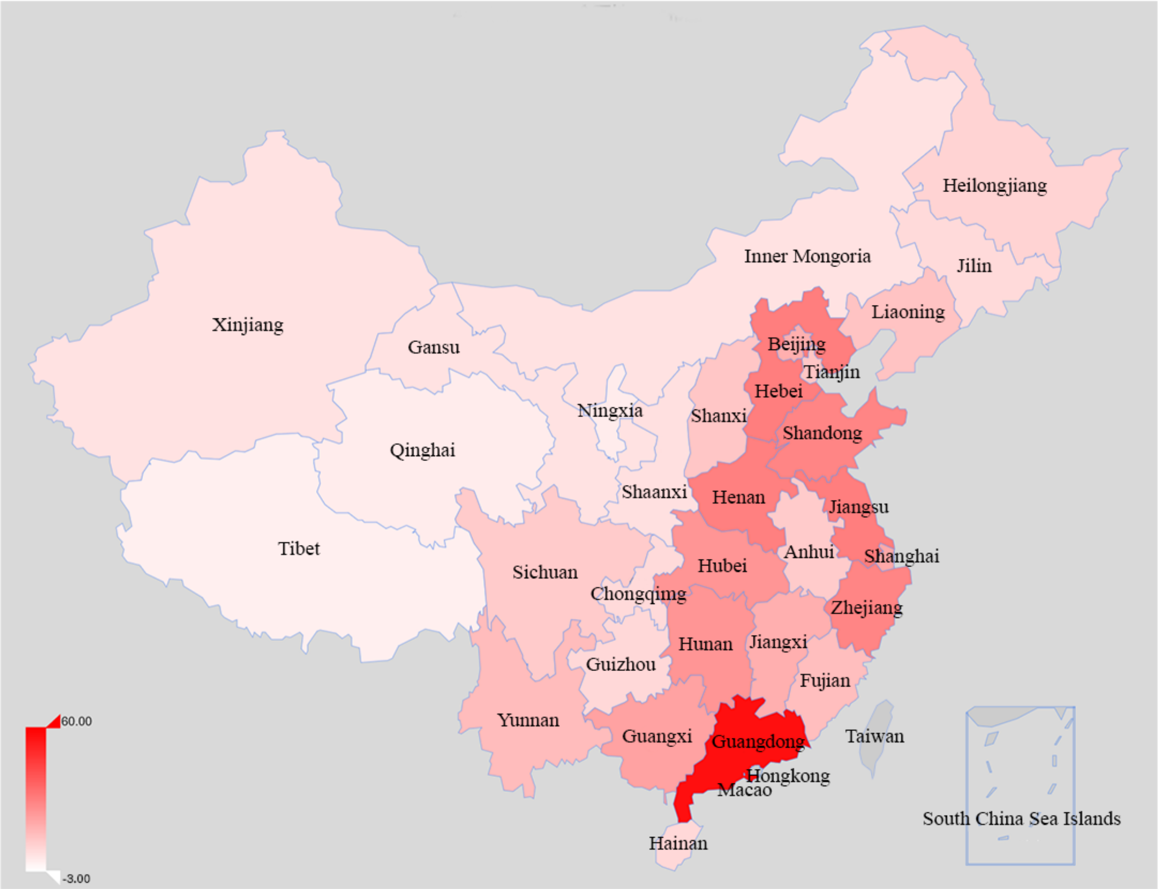


Table 3. The number of accessible technologies (total is the sum of all access technology).

Access to technology	Value, n (%)
Artificial insemination with husband (AIH)	449 (37.02)
Artificial insemination by donor (AID)	60 (5.00)
In vitro fertilization and embryo transfer (IVF-VT)	327 (26.96)
Intracytoplasmic sperm injection (ICSI)	327 (26.96)
Preimplantation genetic diagnosis (PGD)	40 (3.30)

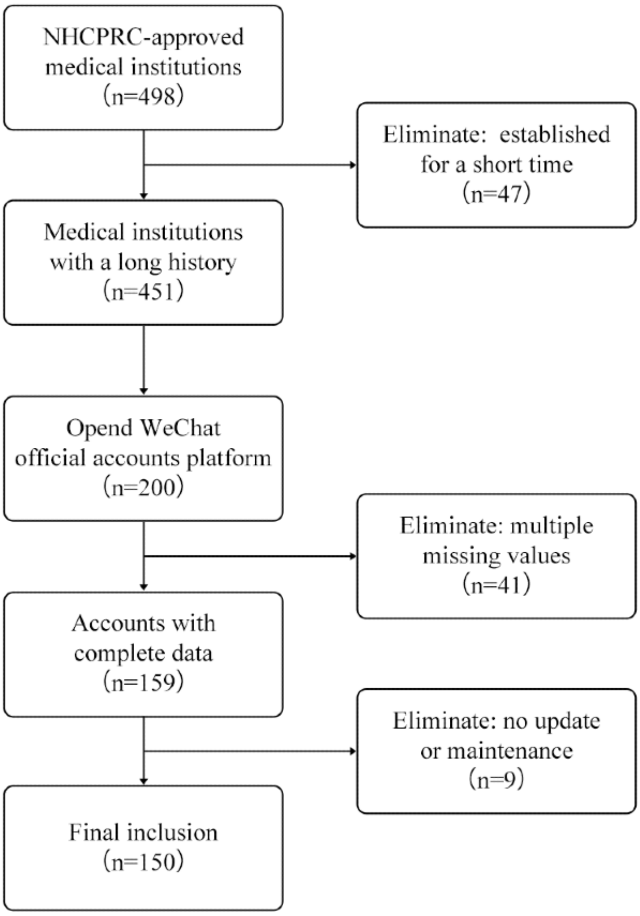
Table 4. The number of ART centers approved by National Health Commission of the People’s Republic of China.

Years	Number, n
2002	12
2004	37
2006	64
2007	95
2011	178
2012	356
2016	451
2018	489

WeChat Accounts Are Uncommon and Their Operation Is Inadequate

We found only 200 ART centers had opened WeChat official accounts, and the operation of these media accounts differed. Of these 200 accounts, 41 had multiple missing values related to their posts, and another nine accounts had not been continuously maintained, updated, or pushed, or only had a few tweets in the past year. Therefore, only 150 accounts ultimately met the criteria for inclusion in the study (Figure 3).

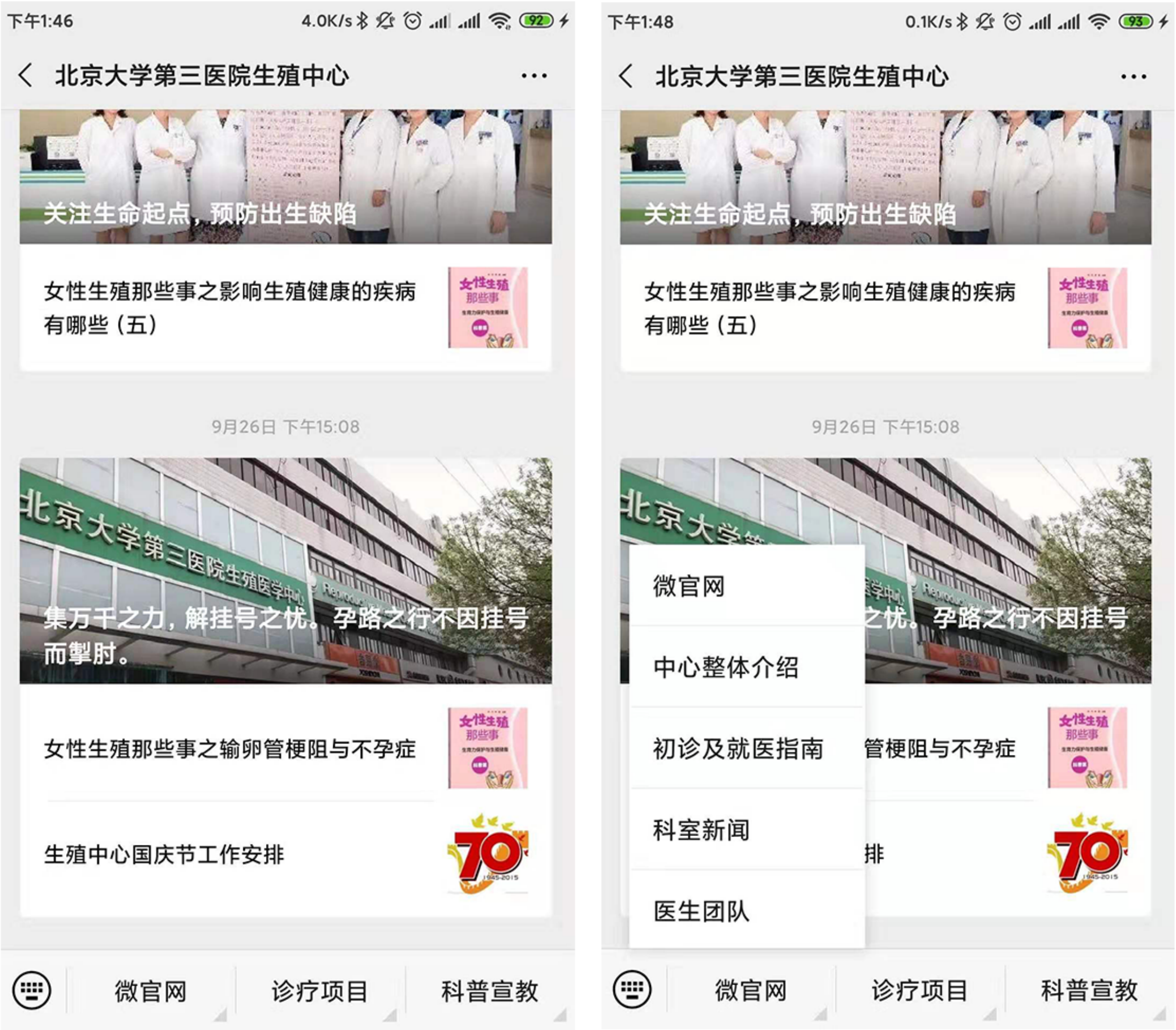
Figure 3. Flowchart of the selection process for WeChat official accounts of assisted reproduction technology centers. NHCPRC: National Health Commission of the People’s Republic of China.



Low Usage and Quality of WeChat Accounts

The main page of a WeChat official account displays posts, including text, graphics, audio, and video. And there is an area at the bottom is called *menu bar*, through which patients can quickly find what they want (Figure 4). On November 14, 2012, Beijing Baodao Healthcare was the first ART institution to open a WeChat official account. However, the emergence of centers using WeChat official accounts did not attract people’s attention, according to our measures (Table 5). The index of the functionality of the WeChat account menu bar showed that the menu bar scores of most accounts were between 5 and 10 points (Table 6).The first wave of ART centers began in 2013 and reached its peak in 2014, indicating that the awareness and utilization of WeChat official accounts by ART centers needs to be further improved in China. The official website and certification of the WeChat official account is an important indicator of patients’ trust. As shown in Tables 7 and 8, there were 127 (63.5%) and 166 (83.0%) accounts with an official website and certification, respectively.

Figure 4. Profile of WeChat official accounts for assisted reproduction technology centers.



北京大学第三医院生殖中心

关注生命起点，预防出生缺陷

女性生殖那些事之影响生殖健康的疾病有哪些（五）

9月26日 下午15:08

集万千之力，解挂号之忧。孕路之行不因挂号而掣肘。

女性生殖那些事之输卵管梗阻与不孕症

生殖中心国庆节工作安排

Table 5. The number of WeChat official accounts foundation date.

Years	Number, n
2012	1
2013	7
2014	50
2015	115
2016	144
2017	183
2018	200

Table 6. Menu bar scores of WeChat official accounts.

Menu bar scores	Number, n
0~2	13
3~4	15
5~6	44
7~8	67
9~10	40
11~12	15
13~14	4
15~18	2

Table 7. The website proportion of WeChat official accounts.

Type	Value, n (%)
With websites	127 (63.5)
Without websites	73 (36.5)

Table 8. The proportion of official hospital certifications.

Type	Value, n (%)
Certified	166 (83.0)
Not certified	34 (17.0)

Extremely Uneven Quantity of Posted Article

As already mentioned, NHCPRC approved 451 medical institutions to conduct human ART, but less than half (200/451, 44.4%) of the ART centers have opened WeChat official accounts, of which 28.4% (128/451) have a subscription account and 16.0% (72/451) have a service account. The overall number of ART centers without an official account is 55.6% (251/451; Multimedia Appendix 1).

In roughly the past year, the cumulative number of posts of these official accounts varied greatly. In ascending order, the medical institutions below the third quartile (Q3) had not posted more than 100 times (Multimedia Appendix 1). Similarly, posting frequency and the number of articles of each post showed the same trend; the Q3 was 5.7 posts per month and 2.9 articles per post (Multimedia Appendix 1).

Afterward, we focused on the number of views of posts. In general, the average number of views of the headlines of the 150 official accounts was 1344.20, whereas the average number of views of the subheadline was only 5.20.

Further analysis of these accounts found that the greatest number of views of the headlines of accounts falling into Q1, Q2, Q3, and Q4 were 375.4, 795.3, 1791.2, and 18445.0, respectively, and the number of views of the subheadline was 128.6, 323.8, 710.0, and 9421.5, respectively (Multimedia Appendix 1). The number of likes was another evaluation index, which measures readers' recognition and approval of tweets. The number of likes for the headline and subheadlines showed trends similar to the number of views (Multimedia Appendix 1).

Ranking of Internet Influence

As mentioned earlier, TOPSIS was used to evaluate the internet's influence for ART centers through their WeChat accounts. To do this, we adopted the objective method to determine the weight of the index (Table 9).

Multimedia Appendix 2 shows the results of the TOPSIS analysis. Among the top 10 medical institutions in the RIIRC are the Reproductive and Genetic Hospital of CITIC-XIANGYA, the Guangdong Institute of Family Planning Science and Technology, The First Affiliated Hospital of Anhui Medical University, the Jiangsu Province Hospital Affiliated with Nanjing Medical University of Medicine, the Peking University Third Hospital, the Kunming IVF Hospital, the Third Affiliated Hospital of Guangzhou Medical University, the Sun Yat-Sen Memorial Hospital of Sun Yat-Sen University, the Northwest Women' and Children's Hospital, and the Reproductive Hospital Affiliated with Shandong University.

Table 9. The index weight of WeChat official accounts for assisted reproduction technology centers.

Index	Weight
x_1	0.029225
x_2	0.027547
x_3	0.067134
x_4	0.090053
x_5	0.027628
x_6	0.042456
x_7	0.038251
x_8	0.06768
x_9	0.058521
x_{10}	0.042434
x_{11}	0.039738
x_{12}	0.047673
x_{13}	0.073306
x_{14}	0.115235
x_{15}	0.099312
x_{16}	0.045186
x_{17}	0.073518
x_{18}	0.008801
x_{19}	0.006303

Discussion

Principal Findings

Taking WeChat official accounts as the breakthrough point, this study analyzed the information about ART centers from the perspective of a semiprofessional patients, using basic information and the number of views and likes to construct a mathematical model and obtained the RIIRC. The major findings involve the geographic distribution of the ART centers, their technical ability, their use of WeChat, and the internet influence for the centers.

Although the 451 ART centers in the sample are distributed throughout the country, their geographic distribution is uneven, and they vary in terms of types of ART they provide. Analysis of the text of posts in their official accounts found the ART centers have insufficient awareness of network exposure and publicity, and some of the centers' internet influence rankings are inconsistent with their medical level and academic status. We also found evidence that internet influence may affect the ART market structure in the future and affect the academic rankings of the medical institutions.

WeChat Official Account Is an Excellent Tool to Popularize Assisted Reproduction Technology Centers

WeChat is an internet instant-messaging social software that was launched by Tencent in January 2011 and launched its official account in July 2012 [19]. As of June 2019, WeChat had 1.08925 billion users who were active every month, and the total number of registered official accounts exceeded 20 million, forming a new type of vast social network, which has increasingly attracted the attention of researchers [20]. In the first half of 2019, the average daily use time of WeChat was 89.8 min, which is 5.1% higher than it was in 2018, making it become one of the apps with the longest daily use time per capita in China [21]. In addition, WeChat has a function that other social software does not have. It is called "Moments (Pengyouquan)," which refers to a circle formed by friends who know each other through WeChat; it is a niche and private circle constructed by an acquaintance relationship chain. This attribute of the "Moments" makes information posted by individuals more likely to be followed and trusted by friends. Given the characteristics of WeChat and WeChat official accounts, we have reasons to believe that WeChat platforms can provide ART centers with a convenient and low-cost opportunity to attract local customers and increase exposure.

The Selected Indices Are Reasonable to Reflect Ranking of Internet Influence on Reproductive Centers

We chose 19 indices as model parameters based on the framework of the WeChat official account, used an objective method to determine the weight of each index, and used TOPSIS to construct a measurement model of internet influence for ART centers (the RIIRC).Some WeChat official account rankings have recently appeared in China, including the WeChat Communication Index and the New Rank Index [16-18]. These rankings and evaluation methods are based on data available from WeChat official accounts, which reasonably reflect the comprehensive internet influence of official accounts to a certain extent and are generally consistent with our methods.

Although the focus of our research was the WeChat official account for ART centers, we also included an index of access to technology. The weight of the access to technology index in our study was 0.006303, which is very low. The reason is that this index assesses the medical level and academic ability of the ART center; it is not a conventional index of internet influence.

The number of views and likes have been used in capacity research as indices to measure the internet influence. We found that the weights for the number of articles of each post (x_5) and posting frequency (x_1) were 0.027628 and 0.029225, respectively, accounting for a relatively low proportion of influence. There was no obvious correlation between these two indices and propagation, indicating that they affect a small portion of the internet influence on the WeChat official accounts, which is consistent with previous results [22].One possible reason is that posting articles too frequently does not increase the average number of views, nor enhance a good reputation. Indeed, frequent posting may even distract patients' attention. The majority of the content posted by the WeChat official accounts of the ART centers was popular scientific articles. The number of views of such articles was acceptable, but the number of likes was not high. Psychological research shows that likes are a way for the public to express their approval of content and ideas [23]. The possible reasons why such articles fail to motivate patients' likes include their scientific nature, the preciseness of posts, and the fitness of the health hotspots. Among the many indices, originality was one of the most important ones for measuring the quality of official accounts and the authors' writing level, and its weight was relatively high (30% of the total weight). Some studies have shown that the average number of views of original articles is 6.54 times greater that of nonoriginal [24], which is consistent with our results. In addition, the data used in our analyses were the average number of views and likes of posts of official accounts during the past year, which avoids the problem of variations in the number of views and likes over shorter time periods. Moreover, the selected indices are data that can be viewed directly by general patients. Assessing the internet behavior of hospitals from the perspective of users can better and more objectively reflect the internet influence of WeChat official accounts for ART centers. In summary, no matter what type, quantity, or weight of an index is used, it can represent the internet influence of the official account for the ART center to a certain extent.

Most Assisted Reproduction Technology Centers' Ranking of Internet Influence on Reproductive Centers Are Inconsistent With Their Medical Level and Academic Status

WeChat launched its official accounts in July 2012, but this tool has not received the attention of ART institutions. By the end of 2018, only 200 ART centers had opened official accounts, and only 150 had operated and maintained them normally. The RIIRC showed that the Reproductive and Genetic Hospital of CITIC-XIANGYA is at the top of the list with 100 points. The Guangdong Institute of Family Planning Science and Technology (66.08), the First Affiliated Hospital of Anhui Medical University (65.86), the Jiangsu Province Hospital Affiliated with Nanjing Medical University of Medicine (50.22), and the Peking University Third Hospital (46.86) are among the top five. At present, the most authoritative list of medical institutions in China is the "Fudan University Hospital Ranking," released in November 2019, which includes rankings of the reproductive medicine specialty [25]. Those rankings focus on discipline structure, clinical technique and medical quality, and the level of scientific research of each hospital. The ART hospitals at the top of these two lists are mainly distributed in the coastal areas of Southeast China and the Bohai Rim region. The ART centers in these areas are rated significantly higher than those of other regions in terms of their opening proportion and internet influence. This may be because of the superior resources, such as open policies, sufficient techniques, and financial support [26,27].

Ranking of Internet Influence on Reproductive Centers May Indicate the Future Market Position of Assisted Reproduction Technology Centers

The Peking University Third Hospital and the Reproductive and Genetic Hospital of CITIC-XIANGYA were the earliest medical institutions to provide ART in China. The first two test-tube babies in mainland China were born in 1988, and these two hospitals were the domestic leaders of ART [28,29]. The ART center of the Peking University Third Hospital has published more than 100 scientific articles in internationally renowned magazines, such as *Nature*, *The Lancet*, *Cell*, and *Proceedings of the National Academy of Sciences*, and it is the most powerful ART center in China [30-33]. The ART cycle of the Reproductive and Genetic Hospital of CITIC-XIANGYA reached 49,000 cases in 2018. In the first quarter of 2019, this increased by 14.84% compared with the same period of last year, making it ART center with the most comprehensive medical technology and the largest number of patients at home and abroad [29]. The Peking University Third Hospital and Reproductive and Genetic Hospital of CITIC-XIANGYA are both at the forefront of the two rankings. They not only lead in terms of scientific research and clinical technique but also attach great importance to internet influence. We have reason to believe that such ART centers can maintain and further enhance their level of medical care, academic status, and internet reputation, so as to expand market advantages and become leaders in ART in China.

The Guangdong Institute of Family Planning Science and Technology, the Third Affiliated Hospital of Guangzhou Medical University, and Sun Yat-Sen Memorial Hospital Sun Yat-Sen University are all located in Guangdong Province of China, ranking 2nd, 7th, and 8th in RIIRC, respectively. The Third Affiliated Hospital of Guangzhou Medical University was ranked 20th in the Fudan University Hospital Ranking, and the other two hospitals were not nominated. There are two possible reasons for the above phenomenon: (1) Guangdong, as the province with the largest gross domestic product in China, has a developed economy and a strong business sense, and it recognizes the importance of public praise and (2) Guangdong Province has a large number of ART centers (as many as 56), and the intensively competitive situation made the ART centers aware of the importance of survival, so they paid attention to their image and reputation on the internet. Jiangsu Province Hospital Affiliated with Nanjing Medical University and The First Affiliated Hospital of Anhui Medical University are ranked 4th and 3rd in RIIRC, respectively, while they rank 9th and 13th in the Fudan University Hospital Rankings, respectively. Both of these ART centers have been established for a long time, have a high level of medical care, and a strong sense of subjective development. With the continuous increase of internet exposure opportunities, the number of patients will definitely increase, which will also improve the clinical technique and medical quality of ART centers. We have reason to believe that the aforementioned ART centers will continue to be important players and the front runners in the ART market in the region by focusing on academic development and online exposure.

We also found another interesting phenomenon, which is that some hospitals are at the top of Fudan University Hospital rankings but in the middle and lower range of the RIIRC. These ART centers are at the forefront in terms of clinical technique and scientific research, but their WeChat official accounts are not well operated, and the number of views is lower. For some public hospitals, though there are still many patients who can help them maintain their market position in the region, the long-term adoption of this approach will lead to the loss of patients, which will lead to their annexation by other public medical institutions or private hospitals

Although the NHCPRC has approved 498 ART centers, the existing medical institutions are far from meeting the needs of the market, and private hospitals and private capital have targeted this market.

Listed company Jinxin Fertility and internet medical platform WeDoctor conducted ART technology services to provide patients with high-quality and personalized treatment programs; its Chengdu Xinan Gynecology Hospital and Kunming IVF Hospital are among the top 10 in China. These private hospitals joined the top 10 in a short period of time; one important reason is that they pay attention to network publicity. Once they have more technology, talent, and capital, they are likely to become leaders in the field of ART, and even surpass the top public medical institutions in the region, because the general public does not know much about the academic standing.

Summary and Limitation

In addition, this study found that in the media era, WeChat can provide hospitals with a convenient and low-cost opportunity to attract customers, as well as provide an easy and efficient way to promote their brand and market. Generally speaking, the following three problems exist.

First, the earliest ART centers using WeChat official accounts were newly established hospitals, rather than older ones, probably because the new hospitals were more open to new technologies. Medical institutions with a long history prefer to focus on patient safety, such as reducing medical malpractice, rather than increasing opportunities to communicate with patients.

Second, in the information age, users have more opportunities to obtain various types of information and less patience for reading. For the WeChat account of an ART center, it is necessary to grasp the frequency, time and user needs of the post, and cultivate the user's website "stickiness" and reading habits through a fixed post mode.

Third, opening and verifying an official account of an ART center on the WeChat app is only the first step for a medical institution to take action. In fact, some hospitals stop updating their accounts after only a few tweets. Although some scholars have proposed detailed guidelines for the use of social media in medical institutions and answering patients' difficult questions, most ART centers do not follow these recommendations [34,35].

The sample size of the ART centers included in this study's sample was limited by two factors. First, although the NHCPRC had approved 498 ART centers by July 4, 2019, 47 new medical institutions were not included in the study because they only had a short history. Other centers had to be excluded because they had incomplete data. Future research should include more ART centers and compare their use of WeChat official accounts and changes over time. Second, all the information used in this study was based on public data, and the cycle number of each ART center has not been reported. The reason for this is that the reporting system of the Chinese Society of reproductive medicine is for internal use only. In addition, the WeChat official accounts of some ART centers were registered in the names of individuals rather than the hospitals or departments, so these official accounts did not fall within the scope of this study.

Although this study has some limitations, we analyzed and compared the operation of WeChat official accounts for ART centers and considered the combined academic influence, which offers guidance for promoting patients' access to these hospitals by strengthening their public image and exposure and a reference point for patients to choose suitable ART medical services. It also analyzed the market situation and future competitive advantage of domestic ART centers and indicated ways in which potential investment institutions can enter the field.

AcknowledgmentsThis work was supported by grants from the National Natural Science Foundation (81774075) of China and Priority Academic Program Development of Jiangsu Higher Education Institutions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The number of views and likes of posts of WeChat official accounts. (a) The proportion of WeChat official accounts. (b) Total times of posts. (c) Posting frequency. (d) The number of articles of each post. (e) Average number of views of headlines. (f) Average number of likes of headlines. (g) Average number of views of subheadlines. (h) Average number of likes of subheadlines.

[PNG File , 97 KB - jmir_v22i6e17997_app1.png]

Multimedia Appendix 2

The ranking of internet influence for assisted reproduction technology centers based on technique for order of preference by similarity to ideal solution.

[DOCX File , 26 KB - jmir_v22i6e17997_app2.docx]

References

1. Junqing W, Qiuying Y, Jianguo T, Wei Y, Liwei B, Yuxian L, et al. Reference value of semen quality in Chinese young men. Contraception 2002 May;65(5):365-368. [doi: 10.1016/s0010-7824(02)00281-0] [Medline: 12057790]
2. Liu J, Larsen U, Wyshak G. Prevalence of primary infertility in China: in-depth analysis of infertility differentials in three minority province/autonomous regions. J Biosoc Sci 2005 Jan;37(1):55-74. [doi: 10.1017/s0021932003006461] [Medline: 15688571]
3. Duron S, Slama R, Ducot B, Bohet A, Sørensen DN, Keiding N, et al. Cumulative incidence rate of medical consultation for fecundity problems--analysis of a prevalent cohort using competing risks. Hum Reprod 2013 Oct;28(10):2872-2879. [doi: 10.1093/humrep/det293] [Medline: 23838160]
4. Schmidt L, Sobotka T, Bentzen JG, Andersen AN, ESHRE ReproductionSociety Task Force. Demographic and medical consequences of the postponement of parenthood. Hum Reprod Update 2012;18(1):29-43. [doi: 10.1093/humupd/dmr040] [Medline: 21989171]
5. Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. Hum Reprod 2007 Jun;22(6):1506-1512. [doi: 10.1093/humrep/dem046] [Medline: 17376819]
6. Bhattacharya S, Porter M, Amalraj E, Templeton A, Hamilton M, Lee AJ, et al. The epidemiology of infertility in the North East of Scotland. Hum Reprod 2009 Dec;24(12):3096-3107. [doi: 10.1093/humrep/dep287] [Medline: 19684046]
7. Slade P, O'Neill C, Simpson AJ, Lashen H. The relationship between perceived stigma, disclosure patterns, support and distress in new attendees at an infertility clinic. Hum Reprod 2007 Aug;22(8):2309-2317. [doi: 10.1093/humrep/dem115] [Medline: 17580298]
8. Bak CW, Seok HH, Song SH, Kim ES, Her YS, Yoon TK. Hormonal imbalances and psychological scars left behind in infertile men. J Androl 2012;33(2):181-189. [doi: 10.2164/jandrol.110.012351] [Medline: 21546616]
9. Merritt MA, de Pari M, Vitonis AF, Titus LJ, Cramer DW, Terry KL. Reproductive characteristics in relation to ovarian cancer risk by histologic pathways. Hum Reprod 2013 May;28(5):1406-1417 [FREE Full text] [doi: 10.1093/humrep/des466] [Medline: 23315066]
10. Jensen TK, Jacobsen R, Christensen K, Nielsen NC, Bostofte E. Good semen quality and life expectancy: a cohort study of 43,277 men. Am J Epidemiol 2009 Sep 1;170(5):559-565. [doi: 10.1093/aje/kwp168] [Medline: 19635736]
11. Wu AK, Elliott P, Katz PP, Smith JF. Time costs of fertility care: the hidden hardship of building a family. Fertil Steril 2013 Jun;99(7):2025-2030 [FREE Full text] [doi: 10.1016/j.fertnstert.2013.01.145] [Medline: 23454007]
12. State Council. China government_central government portal. 2007. [Decision of the State Council on the Cancellation and Adjustment of the Fourth Batch of Administrative Examination and Approval Items] URL: http://www.gov.cn/zwgk/2007-10/12/content_775186.htm [accessed 2007-10-12]
13. National Health Commission of the People's Republic of China. 2019. [List of Medical Institutions Approved to Carry Out Human Assisted Reproductive Technology and Set Up a Human Sperm Bank] URL: http://www.nhc.gov.cn/wjw/fzszjg/201907/bddb71b3de8543f292ba5bbd81c6e750.shtml [accessed 2019-07-05]
14. National Health Commission of the People's Republic of China. 2015. [Interpretation of the Notice of the National Health and Family Planning Commission on Printing and Distributing the Guiding Principles for the Layout Planning of Human Assisted Reproductive Technology (2015 Edition)] URL: http://www.nhc.gov.cn/fys/s3582/201506/197803ad6e8041ba99fb58832787a128.shtml [accessed 2015-05-28]
15. CNNIC. China Internet Network Information Center. 2019. A Statistical Report on the Development of China's Internet URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201908/P020190830356787490958.pdf [accessed 2019-08-30]
16. Qingping L. [GSDATA - Qingbo Big Data]. URL: http://www.gsdata.cn/site/data [accessed 2017-07-01]
17. Xu N. Newrank. 2019. [New Rank Index: NRI] URL: https://www.newrank.cn/public/about/reference.pdf [accessed 2019-01-31]
18. Yueming Y, Pengwei Z. A Method for Evaluating the Influence of the WeChat Public NO. JOURNAL OF INTELLIGENCE 2016;35(9):141-145 [FREE Full text]
19. [WeChat is a Way of Life]. URL: https://weixin.qq.com/ [accessed 2019-09-04]

20. WeChat, CAICT, DCRC. Beijing, China: CAICT; 2019 Mar 04. WeChat employment impact report URL: http://www.caict.ac.cn/kxyj/qwfb/ztbg/201903/P020190304620742061060.pdf [accessed 2020-03-13]
21. QuestMobile. 2018. URL: https://www.questmobile.com.cn/ [accessed 2018-12-01]
22. Jing F, Wei L. A Study on Influential Factors of Wechat Public Accounts Information Transmission Hotness. JOURNAL OF INTELLIGENCE 2016 Feb 18;35(2):157-162 [FREE Full text]
23. Jing L. [Emotions and cultural expressions behind WeChat likes]. New Media Res 2016;2(16):12-13. [doi: 10.3969/j.issn.1671-7597.2016.16.007]
24. Ruoxi Z, Xiaoyan M, Ning L, Lili G, Songlin W, Liang C, et al. Analysis on released messages and its effect on WeChat accounts of public health category. Chinese Journal of health Education 2018;34(9):800-803.
25. Fudan University. 2018. [Chinese Hospital and Specialist Reputation Ranking] URL: http://www.fudanmed.com/institute/news222.aspx [accessed 2019-12-01]
26. Ma L, Xie Q, Shi S, Ye X, Zhao A. Regional maldistribution of China's hospitals based on their structural system. Sustainability 2017;9(6):1046-1063. [doi: 10.3390/su9061046]
27. Zhang T, Xu Y, Ren J, Sun L, Liu C. Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions. Int J Equity Health 2017 Mar 3;16(1):42 [FREE Full text] [doi: 10.1186/s12939-017-0543-9] [Medline: 28253876]
28. Peking University Third Hospital. Department Introduction URL: https://www.puh3.net.cn/szyxzx/ksjj/34623.shtml [accessed 2011-12-12]
29. CITIC Xiangya Reproductive and Genetic Specialty Hospital. URL: http://www.zxxyyy.cn/ [accessed 2019-12-31]
30. Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. Cell 2015 Jun 4;161(6):1437-1452 [FREE Full text] [doi: 10.1016/j.cell.2015.05.015] [Medline: 26046443]
31. Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, et al. The DNA methylation landscape of human early embryos. Nature 2014 Jul 31;511(7511):606-610. [doi: 10.1038/nature13544] [Medline: 25079557]
32. Hou Y, Fan W, Yan L, Li R, Lian Y, Huang J, et al. Genome analyses of single human oocytes. Cell 2013 Dec 19;155(7):1492-1506 [FREE Full text] [doi: 10.1016/j.cell.2013.11.040] [Medline: 24360273]
33. Qiao J, Li R. Fertility preservation: challenges and opportunities. Lancet 2014 Oct 4;384(9950):1246-1247. [doi: 10.1016/S0140-6736(14)61749-9] [Medline: 25283562]
34. Bertot JC, Jaeger PT, Hansen D. The impact of polices on government social media usage: Issues, challenges, and recommendations. Gov Inf Q 2012;29(1):30-40. [doi: 10.1016/j.giq.2011.04.004]
35. Effing R, Spil TA. The social strategy cone: Towards a framework for evaluating social media strategies. Int J Inf Manag 2016 Feb;36(1):1-8. [doi: 10.1016/j.ijinfomgt.2015.07.009]

Abbreviations

ART: assisted reproduction technology
IVF: in vitro fertilization
NHCPRC: National Health Commission of the People's Republic of China
RIIRC: Ranking of Internet Influence on Reproductive Centers
TOPSIS: Technique for Order of Preference by Similarity to Ideal Solution

Edited by G Eysenbach; submitted 28.01.20; peer-reviewed by M Li, J Gou; comments to author 13.02.20; revised version received 20.02.20; accepted 26.02.20; published 10.06.20.

Please cite as:
Shao F, He Z, Zhu Z, Wang X, Zhang J, Shan J, Pan J, Wang H
Internet Influence of Assisted Reproduction Technology Centers in China: Qualitative Study Based on WeChat Official Accounts
J Med Internet Res 2020;22(6):e17997
URL: https://www.jmir.org/2020/6/e17997
doi:10.2196/17997
PMID:32357124

Original Paper

Enhancing Patient Experience With Internet Protocol Addressable Digital Light-Emitting Diode Lighting in Imaging Environments: A Phase I Study

Melanie U Knopp[1,2,3], BS; Katherine Binzel[2], PhD; Chadwick L Wright[2], MD, PhD; Jun Zhang[2], PhD; Michael V Knopp[2], MD, PhD

[1]Department of Sports Medicine, Seaver College, Pepperdine University, Malibu, CA, United States
[2]Department of Radiology, Wright Center of Innovation, The Ohio State University, Columbus, OH, United States
[3]Department of Sport and Health Sciences, Technical University of Munich, Munich, Germany

Corresponding Author:
Michael V Knopp, MD, PhD
Department of Radiology
Wright Center of Innovation
The Ohio State University
395 W 12th Avenue, Suite 430
Columbus, OH, 43210
United States
Phone: 1 6142939998
Email: knopp.16@osu.edu

Abstract

Background: Conventional approaches to improve the quality of clinical patient imaging studies focus predominantly on updating or replacing imaging equipment; however, it is often not considered that patients can also highly influence the diagnostic quality of clinical imaging studies. Patient-specific artifacts can limit the diagnostic image quality, especially when patients are uncomfortable, anxious, or agitated. Imaging facility or environmental conditions can also influence the patient's comfort and willingness to participate in diagnostic imaging studies, especially when performed in visually unesthetic, anxiety-inducing, and technology-intensive imaging centers. When given the opportunity to change a single aspect of the environmental or imaging facility experience, patients feel much more in control of the otherwise unfamiliar and uncomfortable setting. Incorporating commercial, easily adaptable, ambient lighting products within clinical imaging environments allows patients to individually customize their environment for a more personalized and comfortable experience.

Objective: The aim of this pilot study was to use a customizable colored light-emitting diode (LED) lighting system within a clinical imaging environment and demonstrate the feasibility and initial findings of enabling healthy subjects to customize the ambient lighting and color. Improving the patient experience within clinical imaging environments with patient-preferred ambient lighting and color may improve overall patient comfort, compliance, and participation in the imaging study and indirectly contribute to improving diagnostic image quality.

Methods: We installed consumer-based internet protocol addressable LED lights using the ZigBee standard in different imaging rooms within a clinical imaging environment. We recruited healthy volunteers (n=35) to generate pilot data in order to develop a subsequent clinical trial. The visual perception assessment procedure utilized questionnaires with preprogrammed light/color settings and further assessed how subjects preferred ambient light and color within a clinical imaging setting.

Results: Technical implementation using programmable LED lights was performed without any hardware or electrical modifications to the existing clinical imaging environment. Subject testing revealed substantial variabilities in color perception; however, clear trends in subject color preference were noted. In terms of the color hue of the imaging environment, 43% (15/35) found blue and 31% (11/35) found yellow to be the most relaxing. Conversely, 69% (24/35) found red, 17% (6/35) found yellow, and 11% (4/35) found green to be the least relaxing.

Conclusions: With the majority of subjects indicating that colored lighting within a clinical imaging environment would contribute to an improved patient experience, we predict that enabling patients to customize environmental factors like lighting and color to individual preferences will improve patient comfort and patient satisfaction. Improved patient comfort in clinical imaging environments may also help to minimize patient-specific imaging artifacts that can otherwise limit diagnostic image quality.

Trial Registration: ClinicalTrials.gov NCT03456895; https://clinicaltrials.gov/ct2/show/NCT03456895

(J Med Internet Res 2020;22(6):e11839) doi:10.2196/11839

KEYWORDS

ambient lighting; patient comfort; medical imaging; color perception; health care environment; internet protocol–based light-emitting diode lighting

Introduction

To improve the quality of clinical patient imaging studies, imaging environments predominantly consider updating the equipment, and they do not sufficiently appreciate that the quality of an imaging examination is highly influenced by the patient. When there is an unmet need to improve the overall quality of clinical imaging studies, hospitals and imaging centers predominantly focus on updating or replacing the imaging system hardware and rarely consider that patient-specific factors within the imaging environment can also influence overall quality.

When a patient is uncomfortable, anxious, or agitated, any patient motion during the image acquisition can contribute to artifacts that limit the diagnostic quality of the imaging study. This can result in inconclusive results and potentially the need to repeat the imaging study, readminister imaging pharmaceuticals, and re-expose patients to ionizing radiation (eg, radiography, computed tomography [CT], and positron-emission tomography [PET]/CT). Imaging facility or environmental conditions can greatly influence the patient's experience, comfort, and satisfaction with an imaging study.

By providing patients with the ability to choose aspects of their environmental experience, they feel much more in control of an unfamiliar and uncomfortable setting [1]. Researchers used an experimental audio-visual installation in a PET uptake room and reported reduced patient anxiety during the uptake phase prior to ^{18}Fluorine-fluorodeoxyglucose PET imaging [2]. Additionally, studies have reported the use of video goggles as a distraction approach in pediatric patients [3]. The distress experienced in a waiting room was studied, and it set the stage for a more comprehensive approach to address patient comfort and related psychological influences [4]. This study was initiated to enable an in-depth analysis of the perception of light and its psychological influences in order to help develop use case scenarios within imaging environments and provide pilot data for a future prospective clinical trial.

Empirical research indicates that color has a large influence on cognition, affect, and behavior of individuals [5]. Color stimuli consist of multiple dimensions including hue, lightness, and chroma. Hue is defined as the comparability to one of the perceived colors (red, yellow, green, and blue) and is often reported with a hue circle [5]. Lightness is comparable to brightness and is inherently the white to black quality of a color [5]. Chroma resembles saturation and is fundamentally considered the intensity or vividness of the color [6]. Since color is entirely dependent on both photoreceptors and neural processing, it is not a physical quantity, but rather a psychophysical one [7].

In a study associating color and mood, researchers confirmed that color schemes in interior design alone could impact an individual's mood [8]. Another study suggested that generally warm color schemes increase an individual's stimulation and muscle tension, whereas cool color schemes tend to relax and decrease tension [9,10]. However, creating one ideal ambiance applicable to all individuals within an environment may be impractical owing to differing individual characteristics and preferences, suggesting that those environments capable of individually modifying color schemes would be most effective [8].

One way to allow for this flexibility in an environment is to use commercial, easily adaptable, ambient lighting products so that patients can personalize their facility environment experience. With such lighting products available, we embarked upon this feasibility demonstration to explore how consumer products in an imaging environment could be used to achieve better patient comfort. The aim of this pilot study was to use a customizable white/colored light-emitting diode (LED) lighting system within a clinical imaging environment and demonstrate the technical feasibility and initial findings of enabling healthy subjects to individualize the ambient lighting and color.

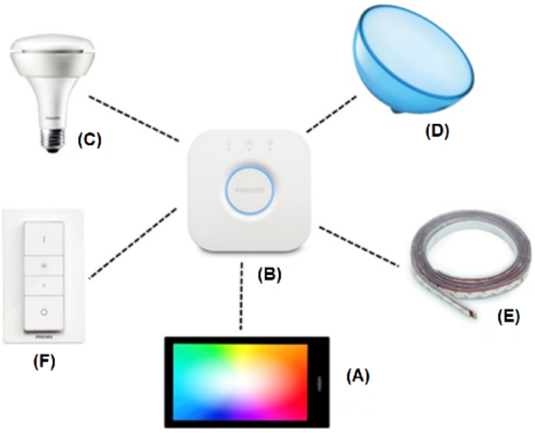
Methods

Internet Protocol Addressable Digital Light-Emitting Diode Lighting

Currently, there are different approaches within the consumer LED lighting market. We chose the ZigBee Alliance standard [11] that supports internet protocol (IP) addressable LED lights and interface devices from different manufacturers. Most of these devices provide dedicated software development tools to support broadly utilizable smart devices. Although various manufacturers have such products, we chose the Philips Hue Personal Wireless Lighting system (Signify, Eindhoven, Netherlands) [12]. This system supports uniquely addressable LED lights that can be remotely controlled and individually programmed to different settings of light hue, chroma, and lightness, making such lights fully interoperable and capable of individualized real-time setting adjustments in imaging environments.

With the lighting system hub in place, we installed BR30 bulbs (Figure 1) [13]. The wirelessly controllable bulbs can be installed in a specific room and then collectively grouped using the unique bulb serial numbers via the Hue smart device app (available on the Apple App Store and Google Play Store). It is also possible to operate multiple hubs on the same network, enabling a well-defined environmental layout with sufficient speed for real-time responsiveness.

Figure 1. A ZigBee internet protocol (IP)-based lighting system and setup with different styles of Philips Hue white and colored lighting bulbs. (A) A smart device via a Wi-Fi router has access to the hub allowing the Philips Hue app to control light settings in various rooms. The screen of the device shows a Hue circle. Figure adapted [13]. (B) The hub (or bridge) is IP connected and communicates with the lights using the ZigBee 2.4 GHz radiofrequency spectrum multihop mesh network to control light (on/off), intensity, and hue. All lights are powered from normal bulb outlets. Each light functions in the multihop setup in both receive and transmit mode. (C) Light-emitting diode (LED) ZigBee bulbs that can be placed in standard light fixtures. (D) Portable LED wireless Philips Hue Go light. (E) Philips Hue Lightstrips can also be added to enhance the room lighting and ambiance. (F) Philips Hue Dimmer Switch can be used to select from preprogramed light color settings. After the color scene is chosen, this device allows the user to both increase and decrease the dim.



Imaging Facilities

We implemented this IP addressable lighting system in three patient preparation rooms, one PET/CT scan room, and one PET/CT control room (Figure 2). Although specialized app designs can be readily achieved using the software development kit (SDK), we focused on the commercially available product app for ease of adoption by others. With the commercially available product app, different ambiences and scenes can be created by individually or collectively setting the hue, chroma, and lightness for each addressable light in the various rooms. In our experiments, each room was individually addressed, but all lights within each room were collectively setup so that each light in the room produced the same hue, chroma, and lightness within the room.

Figure 2. Different color settings in three rooms within an imaging environment. (A) Patient injection/preparation room; (B) scan room of the digital positron-emission tomography/computed tomography (PET/CT) system; and (C) PET/CT control room. In each of these rooms, different color settings are demonstrated as follows: standard bright white light, red light, green light, blue light, and yellow light.

Feasibility Trial Population

For this feasibility phase I trial (ClinicalTrials.gov NCT03456895), we enrolled 35 subjects to participate in the patient injection/preparation room experience only. These subjects did not participate in the PET/CT scan room or PET/CT control room experiences. Each subject was coded with a subject ID number and was evaluated inside a patient injection/preparation room while sitting in a standard patient injection chair with the room door closed. There were several steps.

First, a hardcopy questionnaire covering visual impairment, current mood, favorite color, least favorite color, lighting preferences, and preferred color schemes was completed.

Second, an investigator, who was sitting behind the subjects, set the room's lighting to a full intensity red, identified to the subject as color A. The subjects were exposed to the color A ambience for 30 seconds before being asked to complete another Likert-type scale questionnaire involving five (n=5) or seven (n=30) questions. Upon completion of the color A questionnaire, this procedure was repeated for green light (color B), blue light (color C), and yellow light (color D). The investigator did not name the color but instead used letters to allow the subjects to label the perceived color.

Third, after the four predefined ambient lighting exposures, the subjects completed another survey asking whether they felt that their ability to be productive is influenced by light, which color (A-D) made them feel most relaxed, and which color (A-D) made them feel most anxious.

Fourth, the subjects selected their preferential lighting scheme. The subjects were asked to create their most relaxing light ambience for the room while being able to individually set the hue, chroma, and lightness of the room lights using an app (iOS "Huemote") within a 3-minute timeframe. After the subjects selected their most relaxing light ambience settings, they completed a final questionnaire asking how calm, uncomfortable, and energetic they felt with the individualized relaxing light ambience. The investigator used the "Color Grab" app to identify the subjects' chosen color (specifically, the name of the color and its corresponding red, green, blue [RGB] color codes). The entire testing procedure for the subjects was on average 20 minutes.

For this phase I study, only descriptive statistics are reported. A 5-point Likert-type scale was used to assess subjects' perception with the general layout (1, *strongly disagree*; 2, *disagree*; 3, *neutral*; 4, *agree*; and 5, *strongly agree*). For the charts and overall categorization, we combined in this pilot population all scores of preference involving strong preference and kept the "neutral" answers separate.

Results

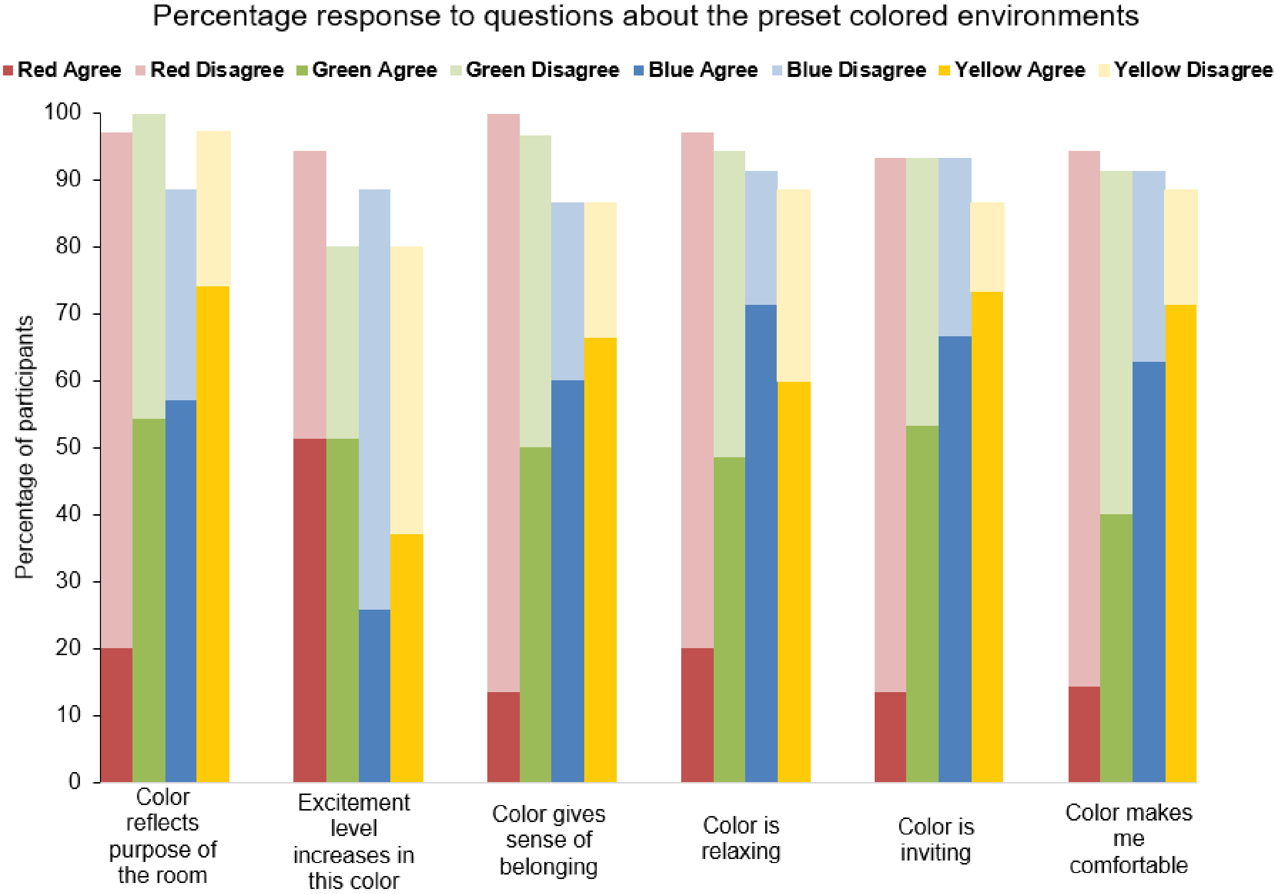
Characterization of the Subjects

Of the 35 subjects, 17 were male and 18 were female. The average age was 33 years (SD 14 years). Among the 35 subjects, 17 wore prescription glasses and seven used contact lenses. The initial questionnaire data revealed substantial variability in perceptions among subjects, with 71% (25/35) choosing natural light as their preferred lighting preference, followed by 17% (6/35) choosing a warm yellow, 6% (2/35) choosing bright white light, and 6% (2/35) choosing very dim light. It should be noted that bright white light is the most common lighting used in health care environments.

Red Lighting (Color A)

Red was the preferred color for 17% (6/35) of subjects and the least preferred for 3% (1/35). In a subset of 30 subjects, red was classified as a warm color by 80% (24/30) of subjects, but was also associated with an increased excitement level, with less favorable ratings for a sense of belonging, being relaxing, being inviting, and being comforting. Only 20% (7/35) of subjects agreed that red was ideal for the patient preparation room, whereas 77% (27/35) disagreed (Figure 3).

Figure 3. Perceptions related to each color. After experiencing each of the four preset colored environments (red, green, blue, and yellow), subjects completed questionnaires on perceptions related to each color. Bars represent the percentage response distributions for the Likert-type preference questions.



Green Lighting (Color B)

Green was the preferred color for 14% (5/35) of subjects and the least preferred for 6% (2/35). In a subset of 30 subjects, green was classified as a warm color by 13% of subjects (4/30), with an almost even split between those who perceived green as relaxing, inviting, comforting, and belonging and those who did not. More than half of the subjects (19/35, 54%) indicated that green was ideal for the patient preparation room (Figure 3).

Blue Lighting (Color C)

Blue was the preferred color for 46% (16/35) of subjects and the least preferred for 6% (2/35). The majority of subjects (22/35, 63%) indicated that blue decreased the excitement level, while increasing the sense of belonging (18/30, 60%), being relaxing (25/35, 71%), being inviting (20/30, 67%), and being comforting (22/35, 63%), despite blue being classified as a warm color by only 13% (4/30). More than half of the subjects (20/35, 57%) reported that blue was ideal for the patient preparation room (Figure 3).

Yellow Lighting (Color D)

Yellow was never the preferred color among subjects and was the least preferred for 20% (7/35) of subjects. In a subset of 30 subjects, yellow was classified as a warm color by 60% (18/30) of subjects and was associated with a decreased excitement level and other favorable ratings for a sense of belonging, being relaxing, being inviting, and being comforting. A high proportion (26/35, 74%) of subjects indicated that yellow was ideal for the patient preparation room (Figure 3).

Combined Color Lighting

The most relaxing color for the patient preparation room was blue in 43% (15/35) of subjects, followed by yellow in 31% (11/35), green in 14% (5/35), and red in 11% (4/35). The most anxiety-inducing color was red in 69% (24/35) of subjects, followed by yellow in 17% (6/35), green in 11% (4/35), and blue in 3% (1/35).

Preferred Lighting

When the subjects individualized their preferred lighting scheme for the patient preparation room and could manually adjust hue, chroma, and lightness, 43% (15/35) preferred blue lighting and 26% (9/35) preferred yellow lighting. The remaining 31% (11/35) of subjects selected a variety of other colors including orange (1/35, 3%), purple (2/35, 6%), red (4/35, 11%), pink (1/35, 3%), and white (3/35, 9%).

Hardware Costs

A large component of implementing this concept is the organizational setup, as the equipment cost is marginal. A starter kit including a hub and three LED IP addressable A19 bulbs costs approximately US $180/€164. An additional hub costs approximately US $60/€55. Extension bulbs, including BR30 bulbs used in this experiment, cost approximately US $50/€45 each. An additional dimmer switch costs approximately US $25/€23.

This feasibility study established the methodology and demonstrated the technical feasibility to test how individual subjects perceive different colored lighting. The technical implementation using consumer-based commercially available LED lights was realized without any hardware or electrical modifications to the existing imaging environment/facility. The installed IP programmable lights were readily controllable using smartphone or tablet-based apps.

Discussion

Principal Findings

This feasibility study demonstrated that a clinical imaging environment can be readily modified to enable an individualized lighting experience with adjustable color schemes using commercially available consumer-based products without specific facility or hardware changes and within manageable costs. The hardware and software used for this study can readily be replaced with alternative technology, making our findings vendor neutral. We have also pilot tested the replacement of hardcopy questionnaires with smartphone or tablet-based surveys that would be even more convenient and efficient for subjects as well as investigators. The subject assessment methodology used in this study can be easily implemented for other clinical applications within health care environments.

Our phase I data indicated that blue lighting was perceived as the most relaxing and was most preferred for the patient preparation room. In contrast, red lighting was perceived as the most anxiety inducing in the patient preparation room. No subject preferred the standard bright white lighting of health care environments, and only 9% (3/35) of subjects preferred a more neutral white lighting for patient preparation rooms when given the opportunity to individualize room lighting.

Available smart device apps can provide patients with individual control of the room’s lighting, allowing them to set their personal preference, which can thereby reduce the stress and anxiety of imaging environments.

Future Studies

Future clinical studies may further examine the impact of ambient light settings on patient experience, mood, perception, and resting physiology. Color psychology is an essential aspect that may explain the variability among subjects’ perception of different lighting schemes. Color has been proven to have a psychological and physical effect on humans, as explained by the Wright theory of color psychology (Table 1) [14]. It was shown that each hue influences particular psychological modes and can therefore affect the mood and behavior of the individual [15].

The psychological properties of the 11 basic colors are described in Table 2. When the human eye sees light, the different wavelengths lead to different perceptions of light. In the retina, these light waves are transformed into electrical impulses that are processed in the thalamus. According to Angela Wright, these psychological colors relate to body, mind, and emotion [16]. The properties further highlight that individuals perceive color differently. Therefore, to maximize comfort levels, lighting systems need to be versatile and readily adaptable to individual user preferences. Our study confirmed that this is achievable using existing consumer LED IP addressable lighting systems.

Given the small number of subjects in this feasibility trial, we are developing future larger clinical trials that will further assess individualized lighting preferences in clinical imaging environments (ie, patient injection/preparation room, PET/CT scan room, and PET/CT control room) and the impact on clinical patients and healthy subjects. Some aspects of future clinical trials will be the randomized assignment of clinical patients and healthy subjects to rooms with standard bright white lighting versus individualized white/colored lighting and the impact on existing institutional patient satisfaction survey scores, as well as net promoter scores. In addition, these trials in clinical patients will evaluate the rates of patient motion artifacts in PET and CT imaging datasets among PET/CT patients imaged under standard lighting versus individualized white/colored lighting in PET/CT scan rooms to address whether individualized ambient lighting reduces patient motion artifacts in diagnostic imaging studies.

Table 1. Seven principal tenants of the Wright theory of color psychology and color harmony [15].

Tenant	Description
1	Each hue affects distinct psychological modes.
2	The psychological effects of color are universal.
3	Every shade, tone, or tint can be classified into one of the four color groups.
4	Every color will harmonize with every other color in the same group.
5	All humanity can be classified into one of four personality types.
6	Each personality type has a natural affinity with one color group.
7	Response to color schemes is influenced by personality type.

Table 2. Current psychological interpretation and association of the 11 basic colors [16].

Color	Psychological aspect	Positive effects	Negative effects
Red	Physical	Physical courage, strength, warmth, energy, basic survival, "fight or flight" stimulation, masculinity, and excitement	Defiance, aggression, visual impact, and strain
Blue	Intellectual	Intelligence, communication, trust, efficiency, serenity, duty, logic, coolness, reflection, and calm	Coldness, aloofness, lack of emotion, and unfriendliness
Yellow	Emotional	Optimism, confidence, self-esteem, extraversion, emotional strength, friendliness, and creativity	Irrationality, fear, emotional fragility, depression, anxiety, and suicide
Green	Balance	Harmony, balance, refreshment, universal love, rest, restoration, reassurance, environmental awareness, equilibrium, and peace	Boredom, stagnation, blandness, and enervation
Violet	Spiritual	Spiritual awareness, containment, vision, luxury, authenticity, truth, and quality	Introversion, decadence, suppression, and inferiority
Orange	N/A[a]	Physical comfort, food, warmth, security, sensuality, passion, abundance, and fun	Deprivation, frustration, frivolity, and immaturity
Pink	N/A	Physical tranquility, nurture, warmth, femininity, love, sexuality, and survival of the species	Inhibition, emotional claustrophobia, emasculation, and physical weakness
Grey	N/A	Psychological neutrality	Lack of confidence, dampness, depression, hibernation, and lack of energy
Black	N/A	Sophistication, glamour, security, emotional safety, efficiency, and substance	Oppression, coldness, menace, and heaviness
White	N/A	Hygiene, sterility, clarity, purity, cleanness, simplicity, sophistication, and efficiency	Sterility, coldness, barriers, unfriendliness, and elitism
Brown	N/A	Seriousness, warmth, nature, earthiness, reliability, and support	Lack of humor, heaviness, and lack of sophistication

[a] N/A: not applicable.

Limitations

A limitation of this phase I trial is that it demonstrates feasibility and findings in healthy subjects with no known diseases in an imaging environment as opposed to patients undergoing imaging procedures for the evaluation of clinical signs or symptoms concerning underlying diseases. Specifically, the healthy subjects in this trial were only tested in the patient preparation rooms, which are used clinically for PET/CT patients, but were not tested in the PET/CT scan room or PET/CT control room. These PET/CT scan room and PET/CT control room evaluations in healthy subjects or clinical patients will be the focus of future studies.

Given that adaptable ambient lighting systems are being incorporated and integrated into many aspects of daily life, we are confident that this work will encourage other researchers to investigate the ambient lighting experience within health care environments in the future.

Comparison With Prior Work

To date, no prior studies have been reported on this topic and the potential utility within clinical imaging environments.

Conclusion

This feasibility study demonstrated that individualizing the ambient lighting experience within a clinical imaging environment can be achieved and implemented using commercially available, consumer-grade, IP addressable lighting products. We found variability among the subjects' perceptions of the various lighting schemes used in this trial, which suggests that an individualized approach to modify the ambient lighting of patient-specific rooms could create more comforting health care environments. This pilot trial found that most subjects (>90%) preferred colored ambient lighting as opposed to standard bright white lighting for patient-specific rooms within the clinical imaging environment.

Acknowledgments

This study was supported by the Ohio Third Frontier TECH 09-028 and the Wright Center of Innovation Development Fund. These funding agencies were not involved in the review or approval of the manuscript. The volunteered time of the study participants is greatly appreciated.

Conflicts of Interest

None declared.

References

1. Mitchell LA, MacDonald RA. An experimental investigation of the effects of preferred and relaxing music listening on pain perception. J Music Ther 2006;43(4):295-316. [doi: 10.1093/jmt/43.4.295] [Medline: 17348757]
2. Vogel WV, Valdés Olmos RA, Tijs TJ, Gillies MF, van Elswijk G, Vogt J. Intervention to lower anxiety of 18F-FDG PET/CT patients by use of audiovisual imagery during the uptake phase before imaging. J Nucl Med Technol 2012 Jun;40(2):92-98 [FREE Full text] [doi: 10.2967/jnmt.111.097964] [Medline: 22570325]
3. Gelfand MJ, Clements C, MacLean JR. Nuclear Medicine Procedures in Children: Special Considerations. Semin Nucl Med 2017 Mar;47(2):110-117. [doi: 10.1053/j.semnuclmed.2016.10.001] [Medline: 28236999]
4. Flory N, Lang EV. Distress in the radiology waiting room. Radiology 2011 Jul;260(1):166-173 [FREE Full text] [doi: 10.1148/radiol.11102211] [Medline: 21474702]
5. Elliot AJ, Maier MA. Color psychology: effects of perceiving color on psychological functioning in humans. Annu Rev Psychol 2014;65:95-120. [doi: 10.1146/annurev-psych-010213-115035] [Medline: 23808916]
6. Landa ER, Fairchild MD. American Scientist. 2005. Charting Colour from the Eye of the Beholder URL: http://code.ucsd.edu/~pcosman/AmSciColor.pdf [accessed 2020-05-20] [WebCite Cache ID 6wh0CqfJ1]
7. Hill GE, McGraw KJ. Bird Colouration: Mechanisms and Measurements. Cambridge, Massachusetts: Harvard University Press; 2006.
8. Kwallek N, Woodson H, Lewis CM, Sales C. Impact of three interior color schemes on worker mood and performance relative to individual environmental sensitivity. Color Res. Appl 1997 Apr;22(2):121-132. [doi: 10.1002/(SICI)1520-6378(199704)22:2<121::AID-COL7>3.0.CO;2-V]
9. Birren F. Colour & human response: Aspect of light and colour bearing on the reactions living things and the welfare of human beings. New York: Van Nostrand Reinhold Co; 1984.
10. Page T, Thorsteinsson G, Ha J. Using Colours to alter Consumer Behaviour and Product Success. International Journal of Contents 2012 Mar 28;8(1):69-73. [doi: 10.5392/ijoc.2012.8.1.069]
11. Zigbee. URL: http://www.zigbee.org [accessed 2020-05-20]
12. Signify: Philips Hue. URL: http://www.meethue.com [accessed 2020-05-20]
13. Indow T, Kanazawa K. Multidimensional mapping of Munsell colors varying in hue, chroma, and value. J Exp Psychol 1960 May;59:330-336. [doi: 10.1037/h0044796] [Medline: 13852802]
14. Wright A. Colour Affects. 2008. Colour Psychology: How It Works URL: http://www.colour-affects.co.uk/how-it-works [accessed 2018-01-23] [WebCite Cache ID 6wh03msvM]
15. Wright A. Colour Affects. 1995. The Wright Theory URL: http://www.colour-affects.co.uk/the-wright-theory [accessed 2018-01-23] [WebCite Cache ID 6wgzpG12e]
16. Wright A. Colour Affects. 2017. Psychological Properties Of Colours URL: http://www.colour-affects.co.uk/psychological-properties-of-colours [accessed 2018-01-23] [WebCite Cache ID 6wgxqrFS6]

Abbreviations

CT: computed tomography
IP: internet protocol
LED: light-emitting diode
PET: positron emission tomography
RGB: red, green, blue

Edited by G Eysenbach; submitted 05.08.18; peer-reviewed by A Mintz, D Du, A Fetit ; comments to author 17.09.18; revised version received 16.09.19; accepted 31.10.19; published 12.06.20.

Please cite as:
Knopp MU, Binzel K, Wright CL, Zhang J, Knopp MV
Enhancing Patient Experience With Internet Protocol Addressable Digital Light-Emitting Diode Lighting in Imaging Environments: A Phase I Study
J Med Internet Res 2020;22(6):e11839
URL: http://www.jmir.org/2020/6/e11839/
doi:10.2196/11839
PMID:32530434

Original Paper

Metrics for Outpatient Portal Use Based on Log File Analysis: Algorithm Development

Gennaro Di Tosto[1], PhD; Ann Scheck McAlearney[1,2,3,4], MS, SCD; Naleef Fareed[1,2], MBA, PhD; Timothy R Huerta[1,2,3,4], MS, PhD

[1] CATALYST: Center for the Advancement of Team Science, Analytics, and Systems Thinking in Health Services and Implementation Science Research, College of Medicine, The Ohio State University, Columbus, OH, United States

[2] Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, United States

[3] Department of Family Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States

[4] Division of Health Services Management and Policy, College of Public Health, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Gennaro Di Tosto, PhD
CATALYST: Center for the Advancement of Team Science, Analytics, and Systems Thinking in Health Services and Implementation Science Research
College of Medicine
The Ohio State University
460 Medical Center Drive, Suite 515
Columbus, OH, 43210
United States
Phone: 1 6142938973
Email: gennaro.ditosto@osumc.edu

Abstract

Background: Web-based outpatient portals help patients engage in the management of their health by allowing them to access their medical information, schedule appointments, track their medications, and communicate with their physicians and care team members. Initial studies have shown that portal adoption positively affects health outcomes; however, early studies typically relied on survey data. Using data from health portal applications, we conducted systematic assessments of patients' use of an outpatient portal to examine how patients engage with the tool.

Objective: This study aimed to document the functionality of an outpatient portal in the context of outpatient care by mining portal usage data and to provide insights into how patients use this tool.

Methods: Using audit log files from the outpatient portal associated with the electronic health record system implemented at a large multihospital academic medical center, we investigated the behavioral traces of a study population of 2607 patients who used the portal between July 2015 and February 2019. Patient portal use was defined as having an active account and having accessed any portal function more than once during the study time frame.

Results: Through our analysis of audit log file data of the number and type of user interactions, we developed a taxonomy of functions and actions and computed analytic metrics, including frequency and comprehensiveness of use. We additionally documented the computational steps required to diagnose artifactual data and arrive at valid usage metrics. Of the 2607 patients in our sample, 2511 were active users of the patients portal where the median number of sessions was 94 (IQR 207). Function use was comprehensive at the patient level, while each session was instead limited to the use of one specific function. Only 17.45% (78,787/451,762) of the sessions were linked to activities involving more than one portal function.

Conclusions: In discussing the full methodological choices made in our analysis, we hope to promote the replicability of our study at other institutions and contribute to the establishment of best practices that can facilitate the adoption of behavioral metrics that enable the measurement of patient engagement based on the outpatient portal use.

(J Med Internet Res 2020;22(6):e16849) doi:10.2196/16849

KEYWORDS

patient portals; health records, personal; health information technology; electronic health record

Introduction

Background

Patient portals are Web-based platforms administered by health care service providers that enable patients to access data held in their personal health record. Meaningful Use (MU), the US federal program that sought the promulgation of electronic health records (EHRs), and its associated incentives resulted in the widespread adoption of patient portals [1].

A notable component of MU required health care organizations to focus on improving patient engagement, specifically through the use of patient portals. This requirement came with the challenge of providing access to vulnerable members of the population (eg, ethnic and racial minorities, persons with disabilities, and those with lower health literacy skills), who are part of groups with higher medical needs in addition to facing technology access issues [2-4]. At the same time research has shown that investments in Web-based portals by health care organizations are increasing in order to meet demands from the general population, as patients want to use these tools at rates faster than they are made available [5]. As a result, many organizations now use outpatient portals that allow patients to access and view their prescriptions and medical information, schedule appointments, and communicate with their doctors [6].

Much of the research assessing the benefits of patient portals has focused on outpatient portals [7]. Studies have shown, for example, that outpatient portal users experience improvements in areas such as reducing risk factors for chronic diseases [8-12] and improving health outcomes [13,14]. However, what characterizes these prior studies is the generalized use of survey data for analysis as opposed to considering direct measures of online actions as a measure of engagement.

One approach to measuring engagement is based on the mining of EHR metadata. Modern EHR systems log every action taken by individuals signed into their systems. These actions create transactional traces on the EHR server in the form of log files, which can be leveraged as the foundation for analysis. Recent studies have made use of log file analysis as a means of assessing and understanding how and why patients engage in their care [7,15,16].

However, barriers remain to using this approach, including difficulty in tracking and accessing the data, the need to process a large volume of information, and the challenge of establishing use metrics. To lower such barriers to adopting this methodology, we add to the existing literature by documenting and describing the data wrangling process and its implications to enable future researchers to apply and validate our approach. To that end, together with our analysis, we publish the Stata code (StataCorp, 2017. Stata Statistical Software: Release 15, StataCorp LLC) required to replicate our analysis to contribute to the creation of shared data models that could enable the quantification of patient portal use behavior across portal types.

Objectives

This methods-focused paper consequently addresses two aims. First, the study aimed to significantly expand the knowledge base of the methodological choices and challenges faced when processing log files. We provide documentation of our approach to process log files using data from our academic health system at the Ohio State University Wexner Medical Center (OSUWMC). This health system offers its patients electronic access to their health care information through the MyChart outpatient portal (Epic Systems Corporation) and has been in use since 2011. We intend to promote a standardized approach to this type of research by including our statistical program files as Multimedia Appendices 1-8 that can be used to analyze log files from Epic Systems. We do so by providing guidance to ensure that future studies adhere to the highest quality of data analysis when using audit log files of outpatient portal use. This approach is similar to the approach undertaken by Huerta et al [17], where a data model and procedure for processing log files from an inpatient portal were provided. However, given the differences between the outpatient portal and the inpatient portal log files, these data are idiosyncratic, affecting how they can be parsed; we note many of these differences in our Discussion section. We address this first study aim below in our Methods section.

The second aim of our study was to provide prescriptive data about outpatient portal usage to demonstrate the implications of our methodological approach and the assumptions about the decisions we made. Our results provide a glimpse into the outpatient portal usage in the context of the care provided at our institution. MyChart portal features include messaging with providers, prescription renewals, appointment information and scheduling, clinical updates, and billing. This approach could be replicated at other institutions that also desire to quantify outpatient portal use for different purposes such as quality improvement (eg, investigate tool performance), increasing patient engagement (eg, identifying the types of users), or for research (eg, summative evaluations of tool uptake). We address this secondary study aim in our Results section, in which we use descriptive statistics to present the portal use data.

Methods

Data Model

This section describes the data model used to examine outpatient portal use. It also documents the data sources required to produce the final data model and lists the recommended computational steps to clean and process the information extracted from the audit log files. The log files audited belong to a sample of 2607 OSUWMC patients, whose patient portal metadata from July 2015 to February 2019 were mined for the purpose of developing the methods to enable the study of patient engagement with Web-based portals.

The analysis of outpatient portal use started with the categorization of all the individual actions performed on the portal by the patients. Actions were then aggregated across functions and over time to quantify user engagement with the technology. The resulting data model is represented in Table 1 and comprises four levels, with each subsequent level comprising the elements of the preceding level.

Table 1. Data model for outpatient portal use.

Data aggregation level	Definition
Action type	Single action performed by the user on the outpatient portal
Portal function	Category grouping user actions under the different functionalities offered through the outpatient portal
Session	Coherent, limited, and uninterrupted use of one or more of the outpatient portal functions via a sequence of user actions
Patient	Use of the outpatient portal, or one of its functions, across the entirety of a patient's recorded sessions

Outpatient portal usage metrics were defined along the following two dimensions:

- Frequency of use: number of times an outpatient portal function has been accessed by a patient
- Comprehensiveness of use: number of unique outpatient portal functions accessed by a patient

Both dimensions were studied at the session and the patient levels of our data model.

Data Sources

We audited the log files from MyChart's instance at our institution to assess portal use. Data from these files included the date and time of the specific actions patients made using the outpatient portal (eg, appointment scheduling and viewing test results).

To support reproducibility of results, we document in Table 2 the sources for the data pertaining to the users' MyChart actions and the status of their accounts by listing the exact variables and tables queried from Epic's Clarity database.

Outpatient portal logs are recorded by Epic in the form of time-stamped sequences of user actions, and an example of this log is reproduced in Table 3. This table contains a sample from the MYC_PT_USER_ACCSS Clarity table, with personal identifiable information redacted. The first three columns present identifiers for a patient's action, which include the medical record number (MRN), a time stamp, and a categorical variable indicating the type of action performed by the patient.

Table 2. Variables queried on the Epic's Clarity database tables, to retrieve data about the users' access to MyChart and their account status history, identified by their master table number by Epic System.

Variable name	Variable description	Master table number
Patient information[a]		
PAT_MRN_ID	Patient identifier	EPT 2061
Account status history log variables[b]		
MYC_STATUS_HX	Update to account status	EPT 28100
MYC_STATUS_TMSTP	Status update time stamp	EPT 28110
MYC_STATUS_MTHD	Technology implementing the update	EPT 28140
MYC_STATUS_CMT	Additional status update description	EPT 28130
User access log variables[c]		
UA_TIME	Action time stamp	WPR 520
MYC_UA_TYPE_C	Action type	WPR 530
UA_EXTENDED_INFO	Additional action type information	WPR 550
UA_SESSION_NUM	Unique session identifier	WPR 561
UA_USER_AGENT	Client user agent information	WPR 566

[a] Variable from the Clarity table: PATIENT.

[b] Variable from the Clarity table: PAT_MYC_STAT_HX.

[c] Variable from the Clarity table: MYC_PT_USER_ACCSS.

Table 3. Example excerpt from MyChart user access logs.

PT_MRN_ID	UA_TIME	MYC_UA_TYPE_C	UA_EXTENDED_INFO	UA_USER_AGENT	UA_SESSION_NUM
xxxxx	2018-09-17 09:20:38	Messaging	Medadvice-form	Null	56xxxx
xxxxx	2018-09-19 12:40:38	Messaging	Medadvice-form	Null	56xxxx
xxxxx	2018-09-19 12:53:12	Messaging	Medadvice-form	Null	56xxxx
xxxxx	2018-09-19 13:47:04	Login	Null	EpicMyChart-iPhone	56xxxx
xxxxx	2018-09-19 13:47:08	Provider List Widget	Get-prov-list	Null	56xxxx
xxxxx	2018-09-19 13:48:05	Messaging	Inbox message list	Null	56xxxx
xxxxx	2018-09-19 13:48:08	Messaging	Message read	Null	56xxxx
xxxxx	2018-09-19 13:48:37	Messaging	Inbox message list	Null	56xxxx
xxxxx	2018-09-19 13:49:05	Messaging	Message read	Null	56xxxx
xxxxx	2018-09-19 13:49:19	Messaging	Message read	Null	56xxxx
xxxxx	2018-09-19 13:50:07	Visits	Get future appt list	Null	56xxxx
xxxxx	2018-09-19 13:50:08	Visits	Get past appt list	Null	56xxxx
xxxxx	2018-09-19 13:50:19	Encounter Details	Past appt dat: 5xxx	Null	56xxxx
xxxxx	2018-09-19 14:06:00	Logout	Logout	Null	56xxxx

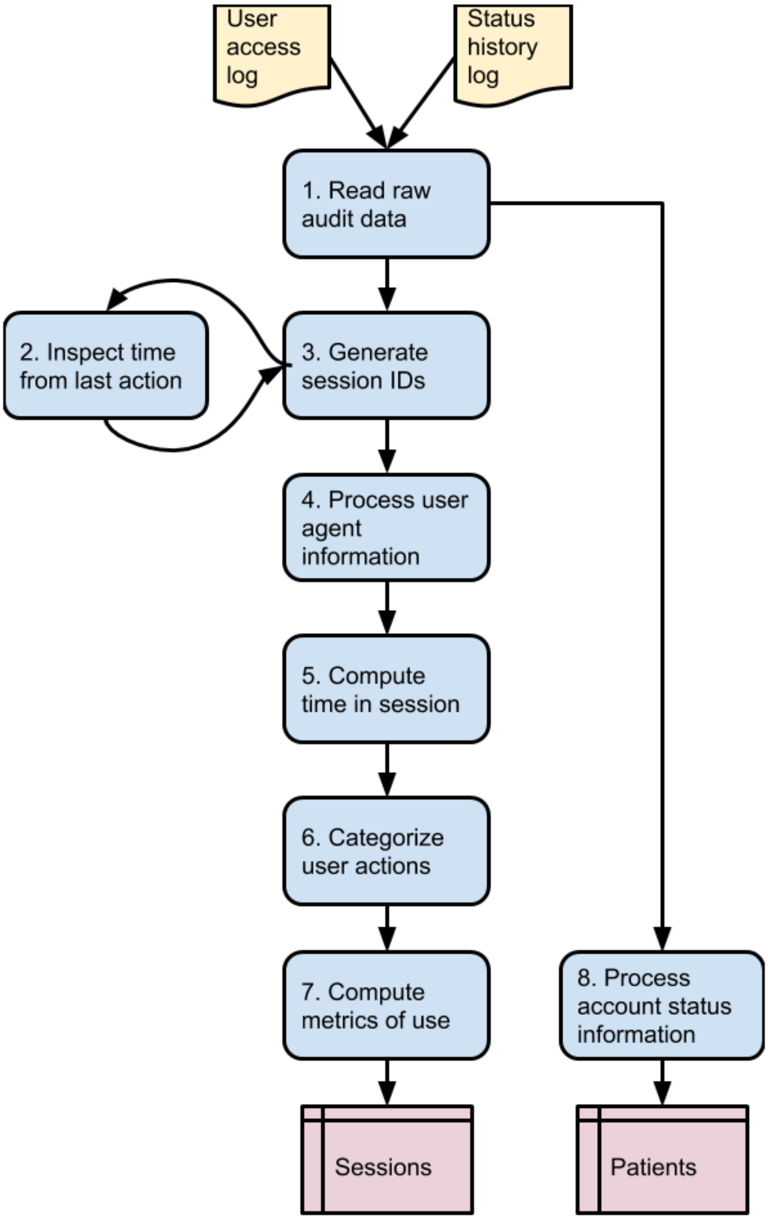
For every recorded user action, Epic's MyChart also reports metadata that can be used to reconstruct the patient's online behavior (ie, what was their activity and how long did it take them to accomplish that action). The variables *UA_EXTENDED_INFO* and *UA_USER_AGENT*, for example, provide important details for the categorization of user actions, necessary for quantifying the use of the different functionalities offered by the outpatient portal. The session number variable *UA_SESSION_NUM*, is instead a unique marker assigned by the server to sequences of user actions that presents patterns of consistent and continuous use.

Data Processing

We next documented the steps taken to process the audit log file data. The result of this data processing is a dataset where each row represents a session and contains information about the time it started, its length, the device type used, and a set of frequency counts for all the activities available to the patient via the outpatient portal.

The steps for processing the raw data were broken down into modules, and these modules are presented in Figure 1 as a flowchart. The assumptions and specific goals of each module are described next. We have included the Stata programming code used to convert the raw data into the final data model as Multimedia Appendices 1-8.

Figure 1. Data processing flowchart.



Module 1: Read Raw Audit Data

In module 1, raw data are imported, and preliminary computations of secondary analytical variables are completed (see Multimedia Appendix 1). In addition to standardizing variables that contain free text or time values, this step generates a categorical variable based on the list of user actions contained in the user access log. This process both establishes MYC_UA_TYPE_C as a sorting variable and controls the ordering of concurrent user actions. For example, when multiple actions are recorded as cooccurrent, log-in attempts are listed first so as to establish a logical order of events.

Module 2: Inspect Time From Last Action

Module 2 quantifies and analyzes the time difference between user actions (see Multimedia Appendix 2). The intent of this step is to control and limit the time gap that identifies a coherent session.

MyChart assigns a *session number* (variable UA_SESSION_NUM) to user actions that signify continuous engagement with the outpatient portal functions; however, this variable can be unreliable. This step enables resolution of issues encountered when working with data from MyChart's Clarity tables. The following are examples of these issues:

- Tracking of the variable might be unavailable: Institutions might deploy a patient portal but postpone tracking of the session number variable. This can result in old audit logs—even for the same patient—that have no server-assigned session information. This is true at our institution for all MyChart data older than May 15, 2016.
- The variable might be tracked intermittently: A session number might fail to be assigned to a subset of user activities, although being recorded correctly for the other activities in the enclosing sequence.
- The variable might be recorded correctly but might indicate implausible sequences of user activities: In some instances, the raw server data can indicate sessions spanning multiple

days (see example in Table 3 above); this is deemed inconsistent with normal user behavior and can be resolved by enforcing a limit to the time gap between consecutive user actions.

Quantifying the time gap between consecutive actions is important as it allows the identification of implausibly long sessions and helps to correct issues with data reliability. This process occurs by calculating a parameter to limit the time gap between user actions, and it has the effect of splitting unusually long sequences into shorter ones, ensuring that the time spans recorded by the audit data measure the actual user engagement with the outpatient portal.

The exact value of the parameter marking the maximum period of inactive time allowed is determined by generating a new variable, *time gap*, which stores the number of seconds separating a user action from the preceding one in the same session and analyzing its density estimation. As shown in Figure 2, this variable appears to have a wide range, with the majority of values at the lower interval, and a long right tail that is approximately log-normal, except for the spike at approximately 20 min. Upon closer inspection, this spike is explained by the time gaps associated with user actions that are recognized as *log-outs* by the MYC_UA_TYPE_C variable. The dotted line reflects the density distribution after this artifact in the data is removed.

Figure 3 shows that establishing a time limit of 1256 seconds (approximately 20 min) allowed us to preserve 99.80% (3,248,044/3,254,561) of the data (excluding the *log-outs*) in their original form.

Figure 2. Density distribution of the time gaps between user actions.

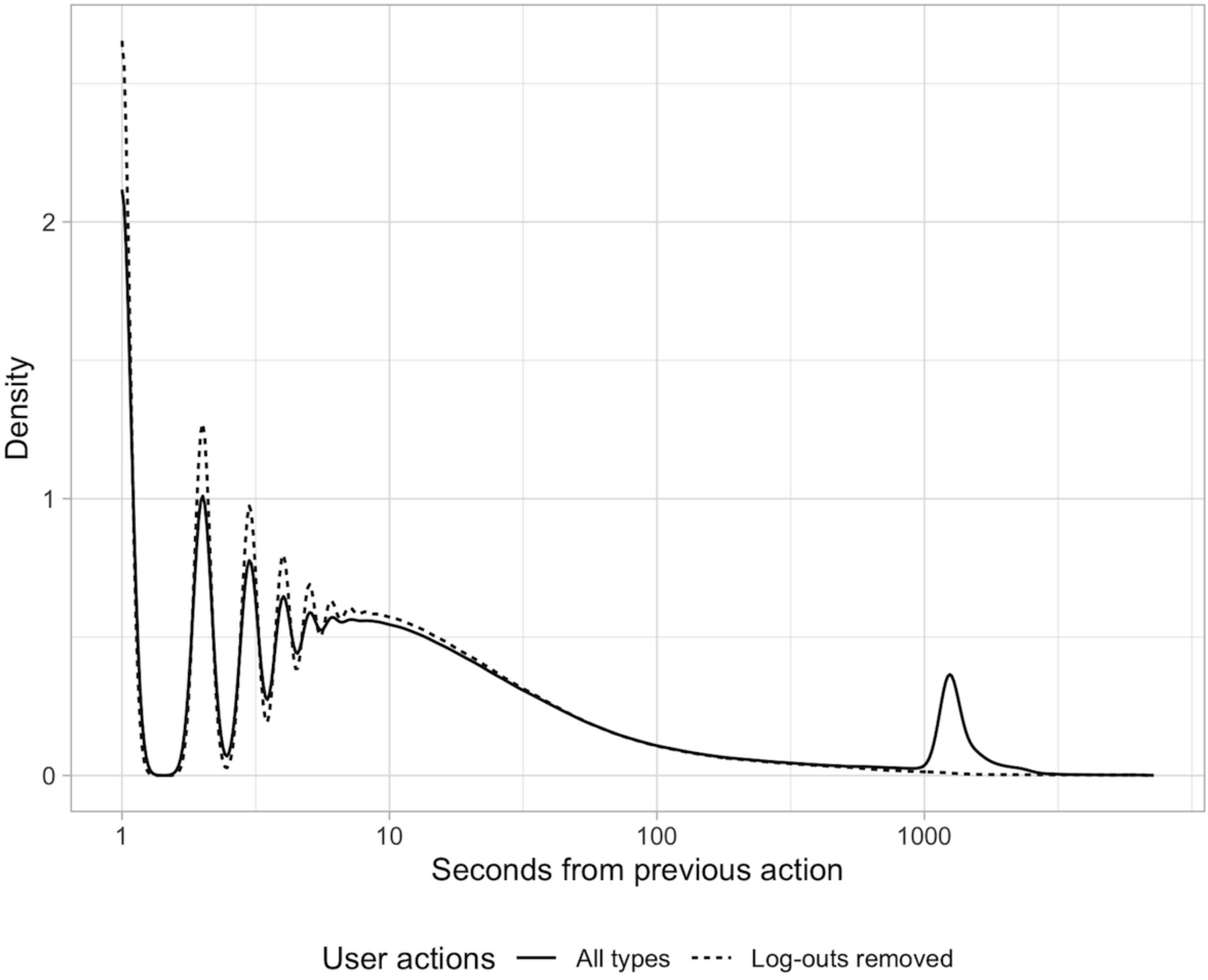
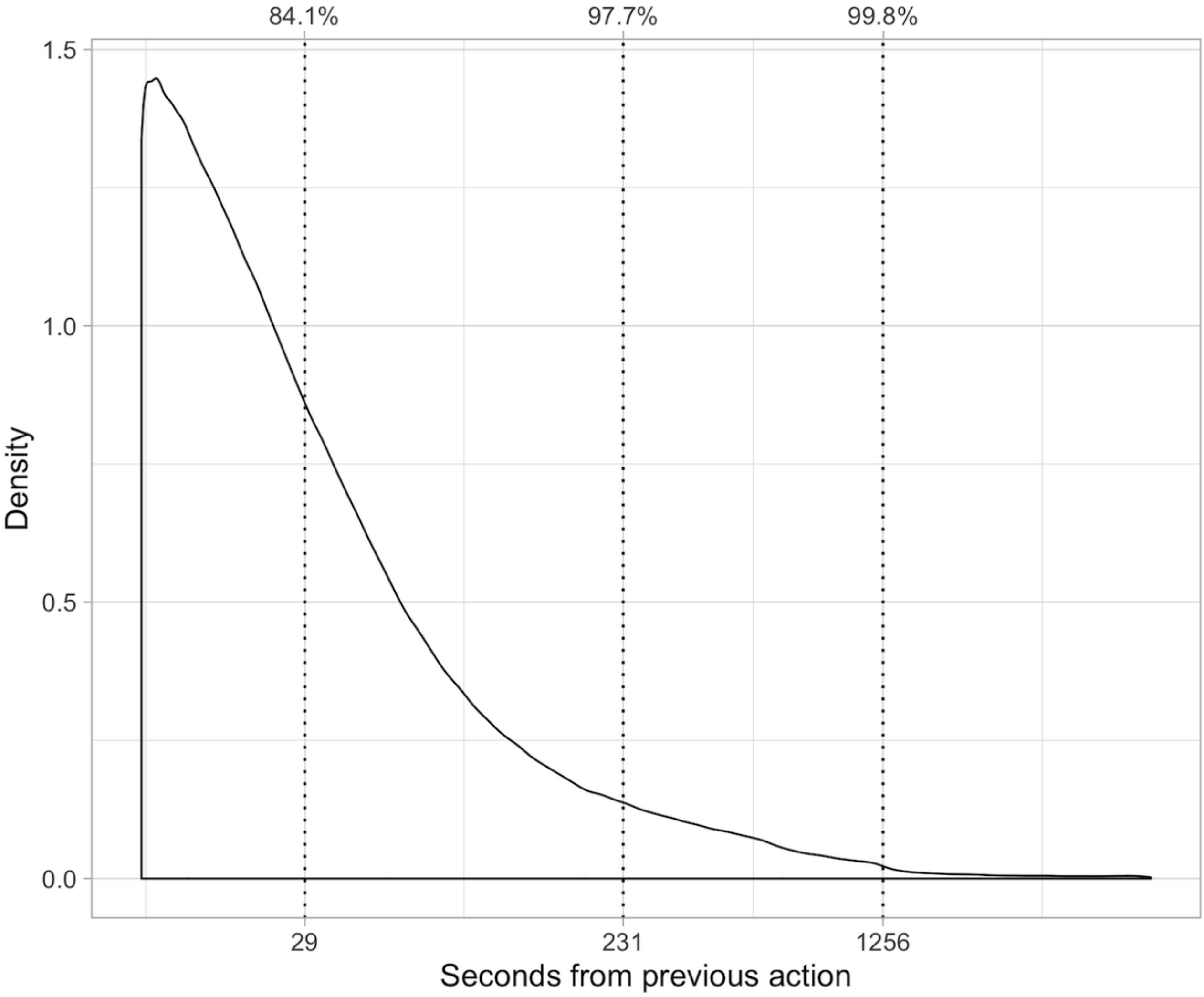


Figure 3. Possible thresholds for inactive time in session. Log-outs action type is removed from the dataset.



Module 3: Generate Session IDs

Module 3 (see Multimedia Appendix 3) creates a session ID that groups together patient activities separated by either a patient's manual log-out or his or her failure to engage with the outpatient portal for what is deemed a significant period of time (defined in module 2). This step supersedes the system-generated UA_SESSION_NUM and groups user activities into sequences of consistent continuous use. By identifying the start and end points of the user sessions, which in most cases would match the user actions marked as *log-in*—for the start point—and as *log-out*—for the end point, an identification label is generated for every user log-in. This label is then extended to all subsequent user-recorded actions, usually up to a recorded log-out. In the event of unusual cases (ie, missing end point, start point, or both), a limit to the time gap between two consecutive user actions is enforced (calculated at approximately 20 min, as described above for module 2). Consequently, all sequences of user actions that are recorded after an inactive time greater than the calculated threshold are regrouped, and a newly created session identification label is generated.

Module 4: Process User Agent Information

User agent information is used in this module to infer the type of device the patient uses to access the outpatient portal (see Mutimedia Appendix 4). Metadata associated with the log-in attempts in the UA_USER_AGENT and UA_EXTENDED_INFO variables can consistently indicate whether access is originating from a desktop or mobile platform. As part of this step, the device type information is stored in a newly created variable, *se_device*, that marks all the actions belonging to the same session.

Module 5: Compute Time in Session

Module 5 calculates the time taken in seconds for each sequence of actions marked with the same session ID (see Multimedia Appendix 5). The algorithm excludes all the *log-outs* from the computation to return an effective time in session. As shown in Figure 3, log-outs consistently overestimate the time a patient remains engaged with the outpatient portal; therefore, we filter out log-outs to avoid artifactual results (see Results section).

Module 6: Categorize User Actions

This module classifies user actions under the categories of the outpatient portal's functions (see Multimedia Appendix 6). The list of labels used to classify user action types, together with their associated grouping functions, are reported in Table 4. This list has been compiled after extensive testing involving the following:

- Controlled interaction with the functionalities of the outpatient portal in a test environment using quality assurance automation tools to verify the traces left by user actions on the server side: As a single user action can be instantiated via multiple system calls—some essential and others ancillary to many other calls, with each one independently leaving traces of its own activation in the audit log—there is no one-to-one correspondence between the user behavior on the client side and the logged portal functions on the server side. We, therefore, undertook a systematic and exhaustive exploration of the functionalities made available to patients via MyChart in a test environment. This test environment mirrored the production portal and included the availability of Clarity tables and automation tools such as Selenium [18], a tool common in quality assurance testing that allows for the reproducibility of workflows and fine-grained control over the timing of actions. This test environment allowed us to generate reproducible patterns in the log data and required the extraction of the signals associated with specific user activities.
- Investigation of more complex cases with information technology (IT) specialists at OSUWMC: In cases where the sole user behavior was not enough to produce unambiguous logs, the help of an IT specialist was enlisted to produce the intended response from the server or to observe in real time if and when fields in the Clarity tables were populated, as a means to verify that the correct data were being collected.
- Clarification of the implementation details with technical support from Epic: In rare cases, information related to implementation details and other clarifications were obtained directly from Epic through their technical support services.

The labels of the action types and portal functions offer a 2-level taxonomy for the categorization of user actions that employ terms from the MyChart user interface. At the higher level, the portal functions are named as follows:

- Messaging: Contains links to the message center, letters to the patient, and an option to request prescription refills
- Visits: List of past and upcoming visits and the ability to schedule and cancel appointments
- My record: List of medications, allergies, medical history and immunization status, health summary and test results, preventive care, and a summary table of the plan of care
- Medical tools: Sharing medical records with others or with other services (Lucy and MyChart Central), participate in research studies, and connect tracking devices (eg, Fitbit)
- Billing: Bills, insurance information, and estimates for common procedures
- Resources: Frequently asked questions, terms and conditions, and patient education
- Proxy: Request proxy status or renew proxy requests
- Preferences: Personal settings, security settings, and notification preferences

Auxiliary action types, such as log-ins, log-outs, and 2-factor authentication, were flagged with a *Miscellaneous* label, but were otherwise excluded from the calculation of portal use metrics. Frequency counts associated with the *Resources* portal function are absent because this set of action types contains links to patient educational information that are all pointers to resources hosted outside MyChart; interaction with these elements, if present, is not captured before the patients navigate away from MyChart.

Table 4. Labels for action types and portal functions used to categorize raw audit log data from Epic's MyChart. For each label, we report the count of sessions in which the action was performed and the percentage represented by the session count over the total number of sessions (N=451,762).

Action type	Portal function	Sessions, n (%)
Letters	Messaging	5318 (1.18)
Message center	Messaging	146,394 (32.41)
Send new message	Messaging	45,391 (10.05)
Appointment details	Visits	17,393 (3.85)
Cancel an appointment	Visits	3665 (0.81)
Driving directions	Visits	232 (0.05)
eCheck-in[a]	Visits	25,246 (5.59)
Schedule an appointment	Visits	67,039 (14.84)
Telemedicine	Visits	13 (0.00)
Upcoming tests	Visits	6784 (1.50)
Allergies	My record	24,345 (5.39)
Current health issues	My record	18,304 (4.05)
Flowsheet	My record	184 (0.04)
Health summary	My record	8989 (1.99)
Immunizations	My record	14,273 (3.16)
Medications	My record	14,450 (3.20)
My conditions	My record	2009 (0.44)
Preventive care	My record	4854 (1.07)
Test results	My record	10,6043 (23.47)
Consolidate EMR[b]	Medical tools	3178 (0.70)
Download my record	Medical tools	53 (0.01)
Research studies	Medical tools	392 (0.09)
Share my record	Medical tools	62 (0.01)
Wallet card	Medical tools	687 (0.15)
Who accessed my record	Medical tools	1267 (0.28)
Bill payment	Billing	6862 (1.52)
Billing account details	Billing	0 (0.00)
Billing account summary	Billing	11432 (2.53)
Change paperless status	Billing	111 (0.02)
Estimates	Billing	838 (0.19)
Insurance summary	Billing	15,538 (3.44)
Update insurance	Billing	11,401 (2.52)
Patient education	Resources	0 (0.00)
Terms and conditions	Resources	0 (0.00)
Proxy forms	Proxy	5763 (1.28)
Proxy renewal request	Proxy	0 (0.00)
Request child proxy access	Proxy	0 (0.00)
Request proxy access	Proxy	84 (0.02)
Switch proxy context	Proxy	556 (0.12)
About me	Preferences	0 (0.00)
Manage my accounts	Preferences	0 (0.00)

Action type	Portal function	Sessions, n (%)
Notifications	Preferences	4349 (0.96)
Personalize	Preferences	1768 (0.39)
Security settings	Preferences	4543 (1.01)

[a] eCheck-in: Electronic check-in.

[b] EMR: electronic medical record.

Module 7: Compute Metrics of Use

The goal of this module is to calculate count values for the portal usage metrics and merge them with the other session-level variables (see Multimedia Appendix 7). The output at the end of this step is a dataset ready for data analysis, organized at the session level. For each constructed session, information about the device used and the time spent in sessions are reported along with frequency counters for the MyChart activities of the taxonomy developed during the previous step. In addition, a reference to the MRN enables the computation of statistics at the patient level as well as the merging of demographic information, when it is available.

Before computing usage metrics, the module filters out redundant records of the same user action from the log file data. Owing to the implementation details affecting users accessing MyChart from a mobile platform, the audit reports can contain many duplicate rows as the same user action can be logged at intervals of 2 seconds until the patient navigates away from the page that is triggering the behavior. Although this does not affect the time in session, it results in inflated action counts for patients accessing the outpatient portal via a mobile phone.

The first part of the code for module 7 controls for the presence of a series of identical user actions, with matching values in both the MYC_UA_TYPE_C and UA_EXTENDED_INFO variables, and deletes all but the first occurrence in each series.

Module 8: Process Account Status Information

The account status information is processed in this module to create a dataset organized at the patient level (see Multimedia Appendix 8). The module accepts as input the list of status changes recorded for each patient's account and processes it to determine the time the patient's account was activated and the current account status. The status of a MyChart account can vary over time, marking, among other things, the sign-up phase (status: *pending activation*), a phase of regular use (status: *active*), and potentially the termination of online use (status: *inactivated*). The resulting dataset, with one row per patient, is suitable to be merged with demographic data to support more detailed analysis of the outpatient portal use.

Results

Session Counts and Activation Status

Using this protocol, we applied the processing steps outlined above to the log files of MyChart activity for the patients in our sample. The following sections present and summarize the results of our analysis of metrics pertaining to the frequency and comprehensiveness of outpatient portal use.

All descriptive statistics were calculated using data from sessions with nonzero duration and from patients who had at least two valid sessions on record. As a result, we discarded all the log-in attempts that did not result in actual user navigation of the outpatient portal; we further refined the dataset to exclude patients who never engaged with MyChart beyond the time required to register an account.

Overall, the dataset contained 451,762 valid sessions (total number of sessions=482,443) for 2511 active patients (total number of patients=2607). On average, each patient had 180 sessions on record; the median number of sessions per patient was 94 sessions, with an IQR of 207 and a maximum recorded count of 6012 (Figure 4).

MyChart account status of the patients included in the final dataset is reported in Table 5, together with the account status of the general hospital population. Not all patient accounts were flagged as *activated*. The majority of *inactivated* accounts were accounts that had been previously active but had been closed because of the death of their owner (*patient deceased*, n=125). The remaining cases represented individuals who never formally concluded the sign-up process or inadvertently invalidated their status after activating their account, but nonetheless engaged with the outpatient portal; as a result, their audit data had been consistently recorded.

The distribution of patients with activated MyChart accounts across age groups (Table 6), although inversely related to the ages of patients, shows that activated accounts remain the largest category across our sample population.

Figure 4. Distribution of session counts per patient.

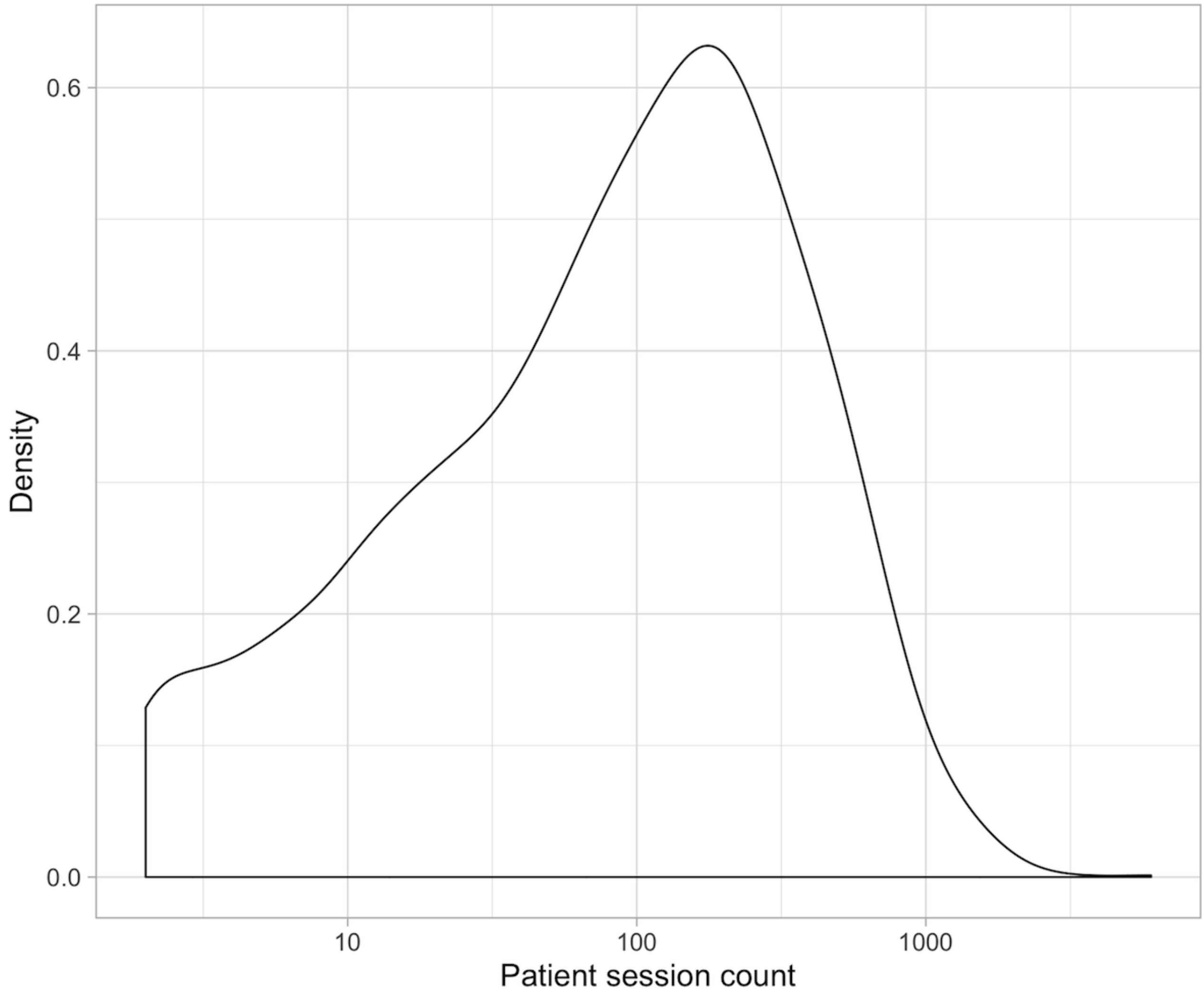


Table 5. Patients' MyChart account status for sampled and overall hospital population.

Account status	Sample population (N=2511), n (%)	Hospital population (N=70,076), n (%)
Activated	2360 (93.99)	39,373 (56.18)
Inactivated	134 (5.34)	12,678 (18.09)
Patient declined	4 (0.16)	6433 (9.18)
Pending activation	13 (0.52)	11,592 (16.54)

Table 6. MyChart account status as a function of patient age group.

Age group (years)	Account status (N=2511), n (%)				
	Activated	Inactivated	Patient declined	Pending activation	Total
18-29	357 (15.13)	7 (5.22)	0 (0.00)	4 (30.77)	368 (14.66)
30-39	468 (19.83)	14 (10.45)	0 (0.00)	2 (15.38)	484 (19.28)
40-49	456 (19.32)	21 (15.67)	1 (25.00)	0 (0.00)	478 (19.04)
50-59	500 (21.19)	36 (26.87)	1 (25.00)	1 (7.69)	538 (21.43)
60-69	429 (18.18)	38 (28.36)	1 (25.00)	3 (23.08)	471 (18.76)
70+	150 (6.36)	18 (13.43)	1 (25.00)	3 (23.08)	172 (6.85)

Frequency of Use

The results in Table 7 show descriptive statistics for the frequency counts associated with each MyChart portal function as defined by the taxonomy of user actions (see also Figures 5 and 6). Of note, at the session level, the median frequency count for the use of all MyChart functions is 0. Although in most cases the ranges are quite broad, the 0 medians suggest that the values are concentrated at lower intervals.

Table 7. Frequency of use at session and patient levels for each of the MyChart portal functions.

Portal function[a]	Session level (N=451,762)				Patient level (N=2511)			
	Minimum value	Median	Maximum value	IQR	Minimum value	Median	Maximum value	IQR
Messaging	0	0	84	1	0	51	2087	134
Visits	0	0	113	0	0	38	1670	108
My record	0	0	227	1	0	87	3503	171
Medical tools	0	0	12	0	0	1	223	3
Billing	0	0	81	0	0	10	659	31
Proxy	0	0	9	0	0	0	498	3
Preferences	0	0	15	0	0	3	228	7

[a] Frequency counts associated with the *Resources* function on the outpatient portal are absent because of technical reasons (see module 6 in the Methods section).

Figure 5. Frequency distribution of the use of the different portal functions across all the sessions in the dataset.

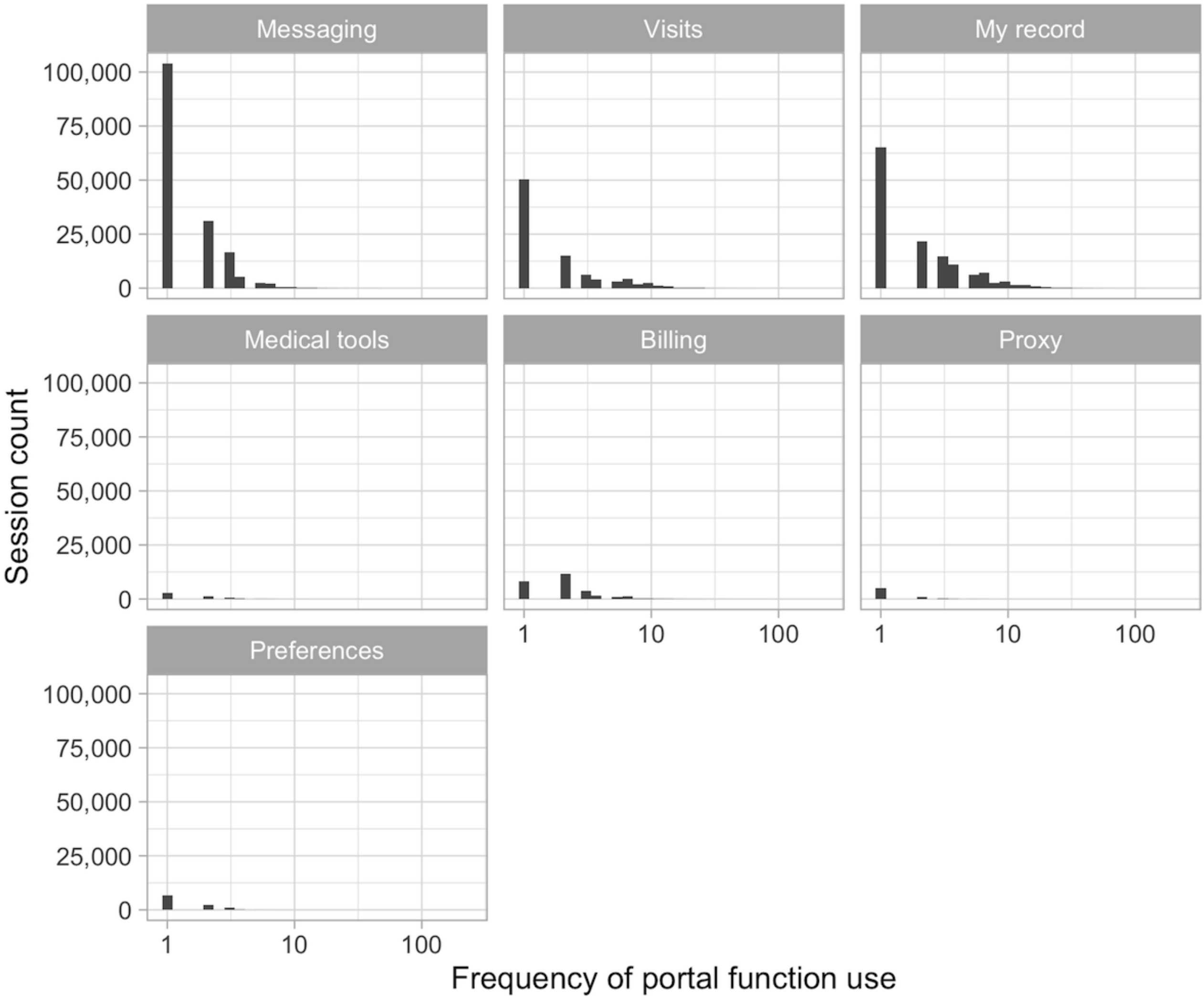
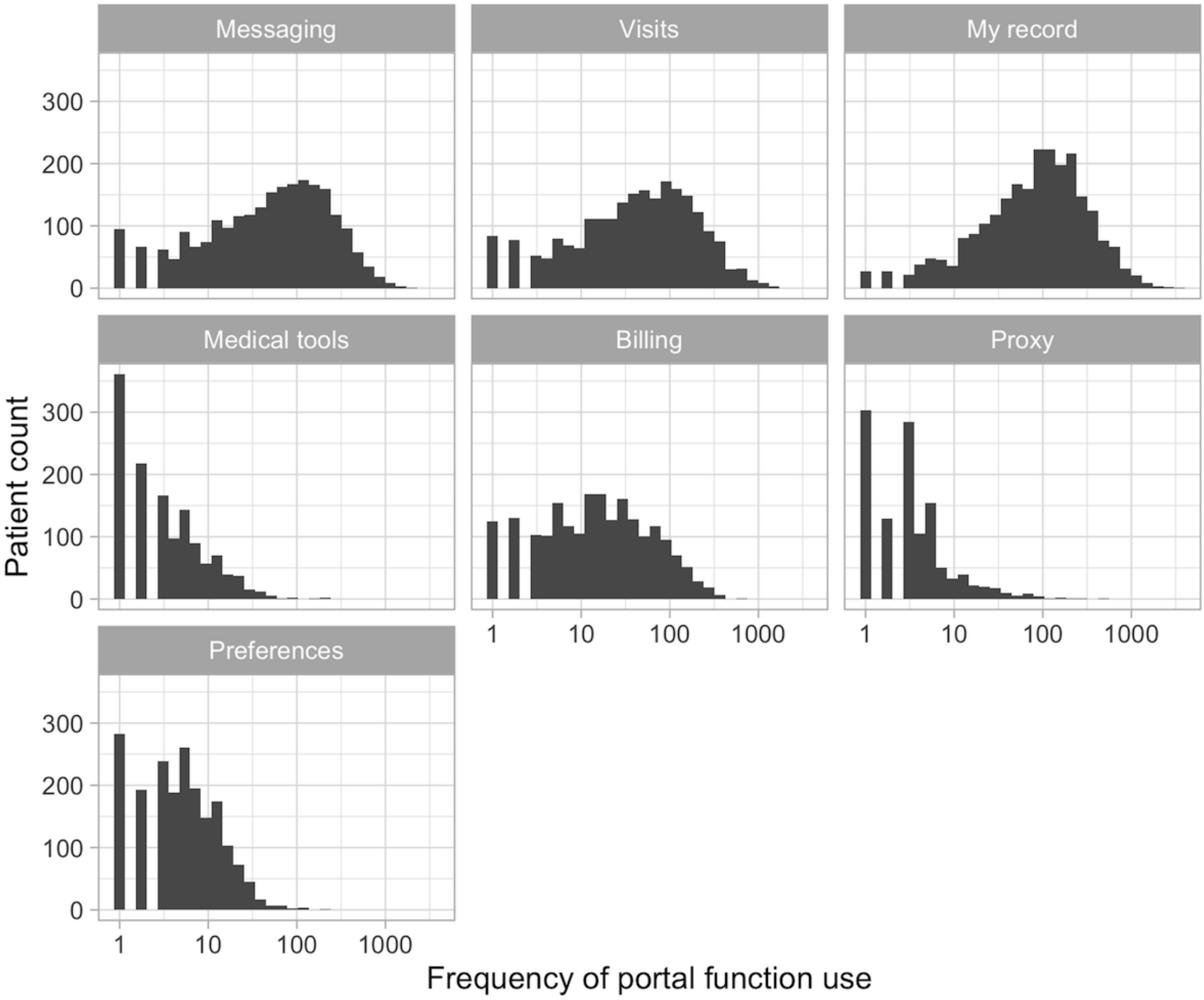


Figure 6. Frequency distribution of the use of the different portal functions across the patient population.



Comprehensiveness of Use

Comprehensiveness of portal function use, defined according to our proposed data model as the number of distinct functions accessed by the patient, provides a summary of how the patient engaged with the outpatient portal.

Comprehensiveness of use results are presented in Table 8. At the session level, in most cases, patients were found to engage with the portal for one specific function; only 17.45% (78,787/451,762) of the sessions were linked to activities involving more than one portal function. At the patient level, most individuals used 3 to 4 of the portal's functions over the course of the study.

Table 8. Comprehensiveness of use.

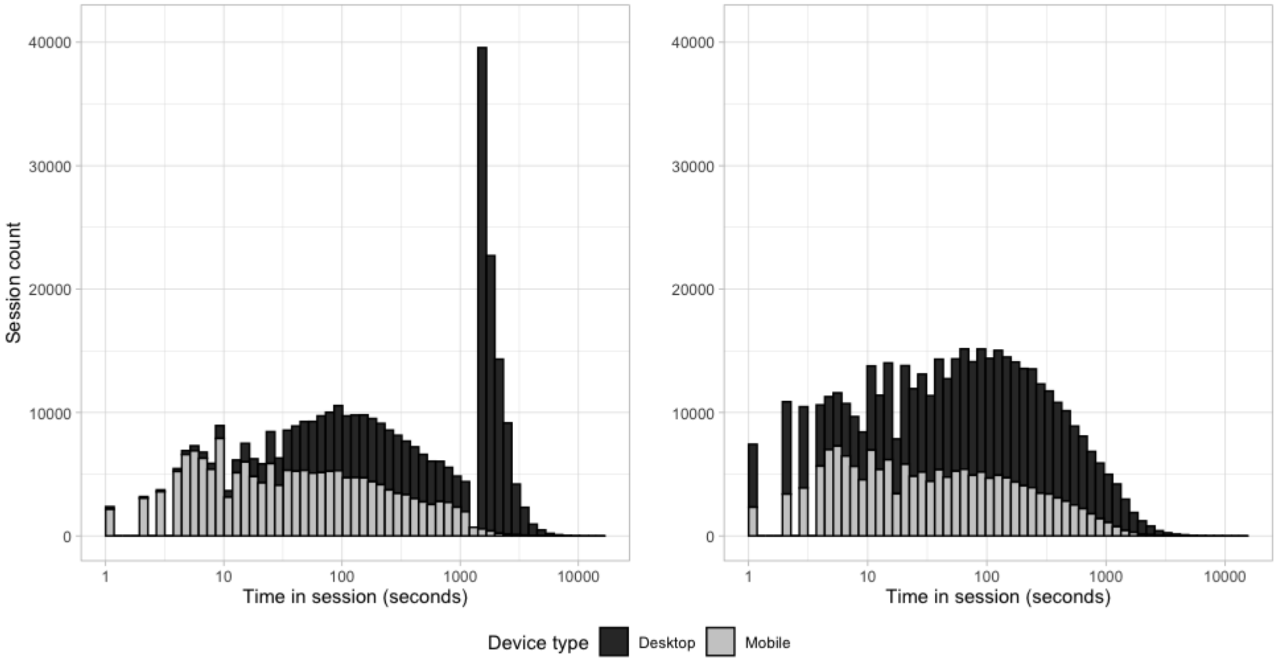
Counts of portal function used	Session (N=451,762), n (%)	Patient (N=2511), n (%)
0	117,819 (26.08)	14 (0.56)
1	255,156 (56.48)	64 (2.55)
2	57,761 (12.79)	274 (10.91)
3	16,215 (3.59)	696 (27.72)
4	4271 (0.95)	1066 (42.45)
5	491 (0.11)	349 (13.90)
6	47 (0.01)	46 (1.83)
7	2 (0.00)	2 (0.08)

Artifacts in the Data

Of note, our data processing methodology also uncovered a series of artifacts in the data, which we were able to correct. The first artifact affected the measurement of the time patients actively engaged with the portal. The distribution of the time spent in sessions calculated on raw server logs showed a bimodal distribution, with most sessions registering a length of a few minutes and a sizable group of longer sessions of approximately 20 min to 22 min (see graph of the left in Figure 7). This second group exclusively comprised patients connecting via a desktop computer (as opposed to a mobile phone), whose connections were kept alive by the server despite the patient being idle on the portal (see description of module 2 in the Methods section). The same analysis performed on audit logs processed with our proposed methodology showed how the distribution of the length of the sessions is instead approximately log-normal (see the graph on the right in Figure 7) and, by getting rid of idle time, better approximated the actual engagement duration of the patient with the portal.

A second type of artifact involved the audit logs mined from the mobile client. As documented in module 7 of the Methods section, duplicate data for MyChart activities recorded during mobile sessions can affect the frequency of use metrics if not properly identified and filtered.

Figure 7. Distribution of time in session by device type. On the left, time in session is calculated on raw server data. On the right, time in session is calculated after processing user access logs, removing log-outs action type, and enforcing a limit of 1256 seconds to the time patients can be idle on the portal.



Discussion

Principal Findings

Patient portals have the potential to improve patient engagement and health outcomes. Our intention with this preliminary study was to uncover the main artifacts in the data and to build a model that starts a conversation on how to use outpatient portal log files for research purposes. As previously noted, evaluations of portal use have, to date, mainly been conducted through the analysis of self-reported behaviors [8]. Furthermore, a review conducted by Irizarry et al [19] about patient portals and patient engagement highlighted the lack of objective measures for usability testing and attributed this gap to the high costs of this type of analysis in terms of time and effort. Our analysis of log files to develop metrics responds to this problem and presents the opportunity to augment self-reports of portal use with the new behavioral measure for portal function use we have described.

Data analytics research has additionally proven effective in describing portal adoption for targeted subpopulations [20]. Our methodology demonstrates how to incorporate information about the status of patient portal accounts in the final dataset (see script for module 8 in the Multimedia Appendix 8). This approach provides information grounded in longitudinal changes to patient accounts that can provide the basis for evaluating portal adoption and use and inform targeted strategies for implementation, improvement, and optimization of use.

Following the recommendations of Sieverink et al [16], our approach highlights the importance of examining outpatient portal use based on how the tool was actually used by patients (ie, including the different functions or elements within it) as opposed to simply reporting the number of log-ins and the frequency of use of the tool. Our results also demonstrate that, based on the level of analysis (ie, patient or session), the patterns of outpatient portal use by function show variability. Interestingly, this variability is more pronounced at the patient level as opposed to the session level. In contrast, at the session level, the usage pattern showed frequent but highly specialized interactions: patients visit the outpatient portal to achieve a single goal and navigate away from it as soon as the goal is achieved. Considering use metrics at the patient level instead, we found that although the portal use spread across all functions, many patients engaged only with a subset of those—a phenomenon that was also detected by Huerta et al [17] in their log file assessment of inpatient portal use.

Although the scope of this descriptive study is limited to the presentation of usage metrics constructed from data elements presently available in our institution's outpatient portal database, the integration of these data with other data sources will present additional opportunities for analysis. Future research on outpatient portal use can explore potential confounders that may influence use, such as demographics, diagnoses, treatments, and patients' communications with care team members. In addition, further work will permit the identification and characterization of user clusters, similar to the work of Jones et al [15] on outpatient portal use among chronic patients and of Fareed et al [21] on inpatient portal use; this approach will improve our understanding of the observed differences in usage patterns at the patient level.

Limitations

We acknowledge two main limitations to this study. First, although we have attempted to generalize our methodology and note procedures that are relevant to log file analysis independent of the institutional context, our study was limited to the analysis of log files from a single outpatient portal system that had retained many of the vendor's technical specifications. For instance, we were unable to obtain log files for functions connected to health educational resources because of the decisions by the EHR vendor and/or our institution to not track such actions. As such information is valuable in measuring patient engagement with their health, additional study is warranted in contexts where these data are collected. Our choice to document the processing steps for our data analysis in the form of Stata do files was motivated by the consideration that enabling reproducible research might bridge the institutional gap and mitigate this limitation. However, applying our methodology in other contexts or with other portal systems will require accounting for the institutional idiosyncrasies and contingencies that influence how the log files are parsed.

Second, with regard to identifying a level of analysis for portal use, as we were working with outpatient records, our data did not have a well-defined cutoff period. This is different from the case of inpatient portal data, where the hospital admission period provides a discrete time frame and level of analysis. We believe that with outpatient portal use, linking portal use to patients' medical history (eg, outpatient portal use during a specific index event such as major surgery) could guide researchers toward more meaningful ways to aggregate log file data and we aim to pursue such an approach in the future.

Conclusions

Although there has been rapid adoption of patient portals across health care organizations, traditional approaches to assessing their impact have not leveraged the potential of detailed audit log file analysis as we have presented. We aimed to model an approach to the outpatient log file analysis by presenting our methodology and describing the challenges associated with accessing and analyzing this type of data. Improving understanding of this process should enable researchers and practitioners to consider this methodology in future studies of outpatient portal use and impact.

Acknowledgments

This work was supported by the Agency for Healthcare Research and Quality grants R01 HS024091-01, R21 HS024349-01, and P30-HS024379.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Stata module 1: read raw audit data.
[TXT File , 2 KB - jmir_v22i6e16849_app1.txt]

Multimedia Appendix 2
Stata module 2: inspect time from last action.
[TXT File , 1 KB - jmir_v22i6e16849_app2.txt]

Multimedia Appendix 3
Stata module 3: generate session ids.
[TXT File , 2 KB - jmir_v22i6e16849_app3.txt]

Multimedia Appendix 4
Stata module 4: process user agent information.
[TXT File , 1 KB - jmir_v22i6e16849_app4.txt]

Multimedia Appendix 5
Stata module 5: compute time in session.
[TXT File , 1 KB - jmir_v22i6e16849_app5.txt]

Multimedia Appendix 6
Stata module 6: categorize user actions.
[TXT File , 37 KB - jmir_v22i6e16849_app6.txt]

Multimedia Appendix 7
Stata module 7: compute metrics of use.
[TXT File , 8 KB - jmir_v22i6e16849_app7.txt]

Multimedia Appendix 8
Stata module 8: process account status information.
[TXT File , 1 KB - jmir_v22i6e16849_app8.txt]

References

1. Henry J, Pylypchuk Y, Patel V. Health IT. 2016 Sep. Electronic Capabilities for Patient Engagement among US Non-Federal Acute Care Hospitals URL: https://www.healthit.gov/sites/default/files/briefs/2015_patient_engagement_data_brief.pdf [accessed 2020-04-01]
2. Chang E, Blondon K, Lyles CR, Jordan L, Ralston JD. Racial/ethnic variation in devices used to access patient portals. Am J Manag Care 2018 Jan 1;24(1):e1-e8 [FREE Full text] [Medline: 29350513]
3. Lyles CR, Fruchterman J, Youdelman M, Schillinger D. Legal, practical, and ethical considerations for making online patient portals accessible for all. Am J Public Health 2017 Oct;107(10):1608-1611. [doi: 10.2105/AJPH.2017.303933] [Medline: 28817324]
4. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, Dabbs AD. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
5. Ford EW, Hesse BW, Huerta TR. Personal health record use in the united states: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
6. Huerta TR, McAlearney AS, Rizer MK. Introducing a patient portal and electronic tablets to inpatient care. Ann Intern Med 2017 Dec 5;167(11):816-817. [doi: 10.7326/M17-1766] [Medline: 29059690]
7. Tsai R, Bell EJ, Woo H, Baldwin K, Pfeffer MA. How patients use a patient portal: an institutional case study of demographics and usage patterns. Appl Clin Inform 2019 Jan;10(1):96-102 [FREE Full text] [doi: 10.1055/s-0038-1677528] [Medline: 30727003]
8. Kelders SM, van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011 Apr 14;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
9. Lau M, Campbell H, Tang T, Thompson DJ, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
10. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011 Sep 30;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
11. Sieverink F, Kelders SM, Braakman-Jansen LM, van Gemert-Pijnen JE. The added value of log file analyses of the use of a personal health record for patients with type 2 diabetes mellitus: preliminary results. J Diabetes Sci Technol 2014 Mar;8(2):247-255 [FREE Full text] [doi: 10.1177/1932296814525696] [Medline: 24876574]
12. van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
13. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014 Oct 24;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]
14. Tulu B, Trudel J, Strong DM, Johnson SA, Sundaresan D, Garber L. Patient portals: an underused resource for improving patient engagement. Chest 2016 Jan;149(1):272-277. [doi: 10.1378/chest.14-2559] [Medline: 26066707]
15. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res 2015 Feb 20;17(2):e42 [FREE Full text] [doi: 10.2196/jmir.3157] [Medline: 25707036]

16. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 7;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
17. Huerta T, Fareed N, Hefner JL, Sieck CJ, Swoboda C, Taylor R, et al. Patient engagement as measured by inpatient portal use: methodology for log file analysis. J Med Internet Res 2019 Mar 25;21(3):e10957 [FREE Full text] [doi: 10.2196/10957] [Medline: 30907733]
18. SeleniumHQ Browser Automation. URL: https://www.selenium.dev/ [accessed 2020-04-02]
19. Irizarry T, Dabbs AD, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
20. Bola J, Charow R, Hope J, Bakas V, Bishop L, Brudnicki S, et al. Adoption strategies for electronic patient portals: employing advanced data mining and analytics. Stud Health Technol Inform 2019;257:36-41. [Medline: 30741169]
21. Fareed N, Walker D, Sieck CJ, Taylor R, Scarborough S, Huerta TR, et al. Inpatient portal clusters: identifying user groups based on portal features. J Am Med Inform Assoc 2019 Jan 1;26(1):28-36. [doi: 10.1093/jamia/ocy147] [Medline: 30476122]

Abbreviations

EHR: electronic health record
IT: information technology
MRN: medical record number
MU: Meaningful Use
OSUWMC: Ohio State University Wexner Medical Center

Edited by G Eysenbach; submitted 30.10.19; peer-reviewed by E Ford, R Bush; comments to author 21.11.19; revised version received 16.12.19; accepted 07.02.20; published 12.06.20.

Please cite as:
Di Tosto G, McAlearney AS, Fareed N, Huerta TR
Metrics for Outpatient Portal Use Based on Log File Analysis: Algorithm Development
J Med Internet Res 2020;22(6):e16849
URL: https://www.jmir.org/2020/6/e16849
doi:10.2196/16849
PMID:32530435

Original Paper

The Impact of Digital-First Consultations on Workload in General Practice: Modeling Study

Chris Salisbury[1], MSc, MD, FRCGP; Mairead Murphy[1], MSc, PhD; Polly Duncan[1], MPH, BMBS, BMedSci

[1]Centre for Academic Primary Care, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Chris Salisbury, MSc, MD, FRCGP
Centre for Academic Primary Care
Department of Population Health Sciences, Bristol Medical School
University of Bristol
Canynge Hall
39 Whatley Road
Bristol, BS8 2PS
United Kingdom
Phone: 44 1173314530
Email: c.salisbury@bristol.ac.uk

Abstract

Background: Health services in many countries are promoting digital-first models of access to general practice based on offering online, video, or telephone consultations before a face-to-face consultation. It is claimed that this will improve access for patients and moderate the workload of doctors. However, improved access could also potentially increase doctors' workload.

Objective: The aim of this study was to explore whether and under what circumstances digital-first access to general practice is likely to decrease or increase general practice workload.

Methods: A process map to delineate primary care access pathways was developed and a model to estimate general practice workload constructed in Microsoft Excel (Microsoft Corp). The model was populated using estimates of key variables obtained from a systematic review of published studies. A MEDLINE search was conducted for studies published in English between January 1, 2000, and September 30, 2019. Included papers provided quantitative data about online, telephone, or video consultations for unselected patients requesting a general practice in-hours consultation for any problem. We excluded studies of general practitioners consulting specialists, consultations not conducted by doctors, and consultations conducted after hours, in secondary care, in specialist services, or for a specific health care problem. Data about the following variables were extracted from the included papers to form the model inputs: the proportion of consultations managed digitally, the proportion of digital consultations completed without a subsequent consultation, the proportion of subsequent consultations conducted by telephone rather than face-to-face, consultation duration, and the proportion of digital consultations that represent new demand. The outcome was general practice workload. The model was used to test the likely impact of different digital-first scenarios, based on the best available evidence and the plausible range of estimates from the published studies. The model allows others to test the impact on workload of varying assumptions about model inputs.

Results: Digital-first approaches are likely to increase general practice workload unless they are shorter, and a higher proportion of patients are managed without a subsequent consultation than observed in most published studies. In our base-case scenarios (based on the best available evidence), digital-first access models using online, telephone, or video consultations are likely to increase general practitioner workload by 25%, 3%, and 31%, respectively. An important determinant of workload is whether the availability of digital-first approaches changes the demand for general practice consultations, but there is little robust evidence to answer this question.

Conclusions: Digital-first approaches to primary care could increase general practice workload unless stringent conditions are met. Justification for these approaches should be based on evidence about the benefits in relation to the costs, rather than assumptions about reductions in workload. Given the potential increase in workload, which in due course could worsen problems of access, these initiatives should be implemented in a staged way alongside careful evaluation.

(J Med Internet Res 2020;22(6):e18203) doi:10.2196/18203

KEYWORDS

general practice; family practice; electronic consultations; remote consultation; telemedicine; telephone consultation; video; access to health care; health care quality, access, and evaluation

Introduction

An increasing number of primary care consultations are being provided under a digital-first model, in which consultations are conducted by telephone, video, email, or online "e-consultation" systems, before offering patients a face-to-face consultation only when necessary. Examples include Doctor On Demand in the United States, Curon in Japan, Ping An Good Doctor in China, and KRY, which operates in several European countries. In England, National Health Service (NHS) policy strongly promotes the use of online consultations [1] and companies such as Babylon GP at Hand, LIVI, and Push Doctor, which offer video consultations free of charge to NHS patients, are expanding rapidly [2]. The introduction of these new access pathways in England will be accelerated by a contract reform framework that will require all general practices to offer online and video consultations by April 2021 and to allow NHS 111 (a national telephone helpline) to directly book face-to-face appointments in local practices [3]. NHS England has recently issued detailed guidance to support the introduction of online consultations [4].

These changes are justified by two main arguments [1-5]. First, they are designed to facilitate quick and convenient access to care by patients, in line with similar changes in how consumers access almost all other services apart from health care. Second, it is argued that the introduction of "online-first" or "telephone-first" models of access will help to manage workload pressures on general practitioners (GPs). However, the twin aims to improve patient access and to manage workload pressures on GPs are likely to be in tension. Whether digital-first models of care decrease or increase general practice workload depends on factors such as the duration of the initial digital consultation and the proportion of these consultations that result in the patient needing a subsequent face-to-face consultation. The impact on general practice workload also depends on whether the demand for consultations is fixed or related to accessibility [6]. At present, many people have difficulty obtaining GP consultations and some may therefore seek help elsewhere or not obtain any professional advice [7]. If quicker and easier access to care means more people contact general practices, this supply-related demand needs to be considered alongside any efficiency gains.

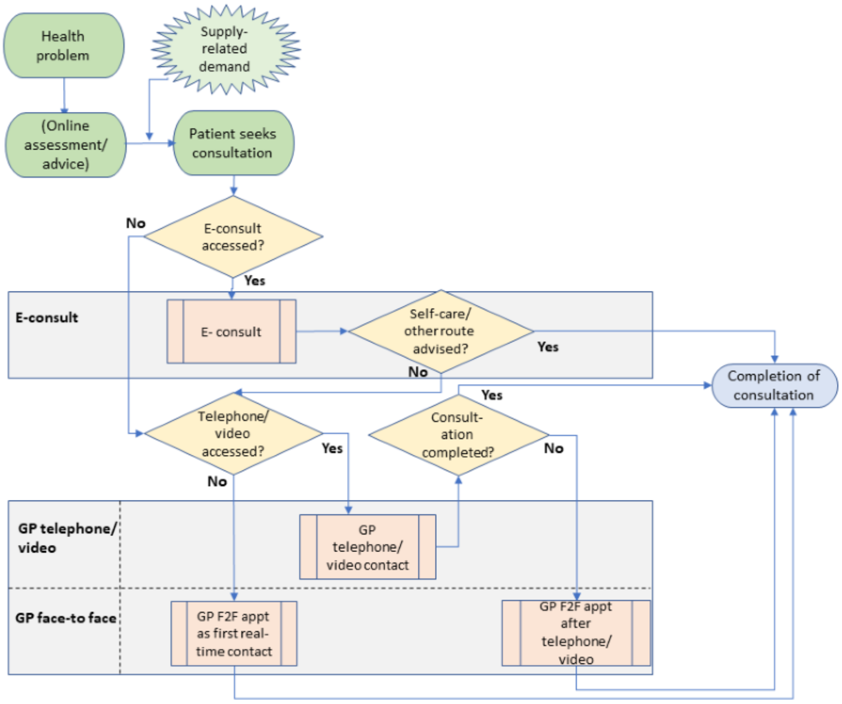
The overall impact on workload in general practice therefore depends on the relationship between several variables. We developed a model to estimate the impact of alternative access pathways on general practice workload and populated the model using a systematic review of studies of digital consultations in primary care. The aim of this study was to inform debate about to what extent, and under what circumstances, digital-first primary care consultations are likely to decrease or increase general practice workload.

Methods

Overview

We developed a process map to delineate the access pathway from when patients first seek a general practice consultation through to obtaining definitive assessment and care (Figure 1). Given the interconnectedness of health care systems, any such model is a simplification and must have a defined scope. Our process map begins with a patient having a health problem and considering requesting a GP consultation. They will often seek advice, which may be from family, a pharmacist, or online through an internet search, an automated symptom-checker app, or a patient forum [7]. Our model begins at the point following this, when a patient actively contacts a general practice requesting a consultation. It therefore excludes administrative issues conventionally dealt with by receptionists (for example, repeat prescriptions), consultations usually managed by nurses (for example, vaccinations), and patients who complete an online application but do not seek contact with a GP. The model includes consultations directly necessitated by a previous step (for example, a face-to-face consultation resulting from an online consultation) but ignores follow-up consultations. The model is limited to the impact on GPs, although we recognize that changes in access to GPs can have consequences for administrative and nursing staff and for demand on other parts of the health service.

Figure 1. Process map: impact of digital first access pathways on general practitioner (GP) workload. E-consult: online consultation. F2F: face-to-face.



We populated the workload model using data from a systematic review for 3 scenarios, representing the following 3 access pathways:

- In an online (or e-consultation) first model, patients describe their problem using an electronic form that may involve an automated algorithm or a less structured form. Administrative requests are dealt with by receptionists (excluded by our model). Requests for medical advice are reviewed by a GP who responds with a message or prescription, or a telephone, video, or face-to-face appointment.
- In a telephone-first model, the GP attempts to resolve the problem by telephone if possible, but if not, invites the patient to a face-to-face consultation.
- A video-first model follows a similar pattern, but there may be differences in variables such as the length of the consultation and the proportion of contacts that require a subsequent face-to-face consultation.

Model

The outcome for the model is percentage change in GP workload using the digital approach compared with GP workload using a conventional approach, in which most patients have face-to-face consultations, but a small proportion have telephone consultations. We created a dynamic spreadsheet model that allowed us to calculate how GP workload changes depending on the values of the key variables in the process map (such as the duration of digital consultations, or supply-related changes in demand). The results from our model are expressed as a percentage, so a result of 10% would mean 10% more hours of work than under conventional care.

The estimates for the conventional approach came from a large and rigorous study of GP consultations in England [8]. The study showed that 86% of consultations were conducted face-to-face and 14% were conducted by telephone, with a mean duration of 9 minutes and 5 minutes, respectively. The estimates for the new digital approaches came from a systematic review, described below. The results from the model illustrate the impact of a central base-case estimate, based on the best available evidence from our systematic review. For each variable in the model, we also considered plausible upper and lower limit estimates based on outlier studies from the review or on our own informed opinion in the absence of evidence. We have made the dynamic model available online, so that readers can test the effect of their own assumptions and estimates [9].

Systematic Review

To include evidence-based estimates in the model, we conducted a systematic review to identify studies of any design that provided quantitative data about digital consultations in primary care, including consultations by telephone, email, e-consultation systems, or video. The quantitative variables of interest were the proportion of consultations managed digitally, completion rate (digital consultations completed without needing a subsequent GP consultation), the proportion of subsequent consultations managed by telephone rather than face-to-face following an online consultation, the duration of different types of consultation, and any indicators of changes in demand or workload after the introduction of a digital-first model.

A change in the number of requests for health care following improved access is commonly referred to as supply-induced demand [6], which implies that a change in services has caused increased demand. However, in this study we used the more neutral term “supply-related demand” because it could equally represent underlying demand, which becomes visible only once access is improved. We sought to estimate the proportion of digital consultations that represented new supply-related demand based on (per month) the number of contacts of all types after introducing digital consultations, minus the number of contacts before introducing digital consultations, minus the number of duplicate consultations where patients had a face-to-face consultation directly resulting from a digital consultation, divided by the number of digital consultations.

Our focus was on consultations for undifferentiated problems between a patient and a GP; therefore, we excluded studies of communication between health professionals (for example, GPs consulting specialists), after-hours consultations, specialist or secondary care consultations, consultations not conducted by GPs, studies limited to a specific type of health problem, qualitative studies, studies of patient or GP opinion, and systematic reviews that did not provide any new quantitative analyses beyond the already-included papers. Our search strategy included terms for general practice, family practice, or primary care, or papers published in leading primary care journals, combined with a wide range of terms relating to telephone, online, digital, or video consultations. We included papers identified through the bibliographies of other papers. Multimedia Appendix 1 shows the search strategy, which was conducted in MEDLINE. We restricted searches to papers published since January 1, 2000, in English, and in developed countries to focus on papers of current relevance to the United Kingdom and other similar health care systems. We did not attempt to grade the quality of the studies or assess the risk of bias. The searches were updated to September 30, 2019. CS and PD reviewed the titles and abstracts of papers independently, retaining any papers that were potentially relevant. They then independently reviewed the full text of these papers to identify those that met the inclusion criteria. CS extracted quantitative data about the variables of interest, and this was checked by PD. Disagreements between reviewers were resolved by discussion involving the third author, MM.

Results

The systematic review identified 1246 papers, of which 90 were judged to be potentially relevant based on their titles and abstracts (Multimedia Appendix 2). Of these, 29 papers provided data of relevance to this study (Multimedia Appendix 3) [8,10-37]. Table 1 shows the estimates for the variables included in the workload model.

Based on our workload model, the final row in Table 1 shows workload in general practice using our central estimates for each scenario, compared with a conventional pathway based predominantly on face-to-face consultations.

The dynamic model makes it possible to test the sensitivity of the model to different assumptions by graphically showing the impact on workload (y-axis) of changing any 2 variables simultaneously (x-axis and legend). In Figures 2 and 3, we show 2 different scenarios as examples; readers can use the model to test their own scenarios [9]. Figure 2 shows that if the average duration of a telephone call is 5 minutes, a telephone-first approach has the potential to reduce workload if at least 55% of telephone consultations are completed without needing a subsequent face-to-face consultation, assuming no supply-related increase in demand. Figure 3 demonstrates that an online-first approach could reduce GP workload only if there is minimal increased demand (<4% of online contacts represent new demand) and about 50% of all requests are resolved in one online contact, but it has the potential to substantially increase workload if there is any supply-related increase in demand. However, 2 recent UK studies suggest that only about 30% of online consultation requests are resolved entirely online [13,17], and several studies suggest that there could be a substantial inflation in demand [14,16,26,31].

Table 1. Alternatives to face-to-face consultation: default values for variables in workload model. All values can be altered in the interactive model to test different scenarios and assumptions. Values without citations are authors' estimates.

Variable	E-consultation[a]			Telephone			Video		
	Base case	Lower value	Upper value	Base case	Lower value	Upper value	Base case	Lower value	Upper value
Access rate: Consultations initially requested in this way[b] (%)	90[b]	0.01 [17]	100	93 [12]	10 [20]	100	90[b]	50	100
Completion rate									
Digital consultations completed without needing a subsequent consultation (%)	30 [17]	28 [13]	70 [33][c]	52 [25,29]	40	90 [20]	65	50	83 [10]
Of those having a real-time consultation after e-consultation, consultations having phone rather than face-to-face consultation (%)	46 [17]	20	90	N/A[d]	N/A	N/A	N/A	N/A	N/A
Average time: Average time spent by GP[e] on this type of consultation (minutes)	4	3	5 [17]	5 [8]	4 [11,15]	6 [29]	9[f]	6 [21]	15 [34]
Supply-related demand: Alternative form consultations that are new demand (%)	10	–10	30	0 [29]	–10	30	0	–10	30
Total workload resource compared with conventional care, using base case assumptions (%)	25	N/A	N/A	3	N/A	N/A	31	N/A	N/A

[a] E-consultation: online consultation.

[b] At present, usage of e-consultation and video consultation in the United Kingdom is generally very low, so the impact is minimal. For the base cases, we have modeled a scenario in which the use of these alternatives is usual.

[c] Penza et al report a 66% completion rate [33]. Completion rates of 70% are claimed by eConsult, cited by Marshall et al [2]. Longman reports similar experiences in practices using askmyGP [38].

[d] N/A: Not applicable.

[e] GP: general practitioner.

[f] Assumed to be similar to conventional face-to-face care.

Figure 2. Impact of telephone consultations on general practitioner workload: varying telephone completion rate and call duration.

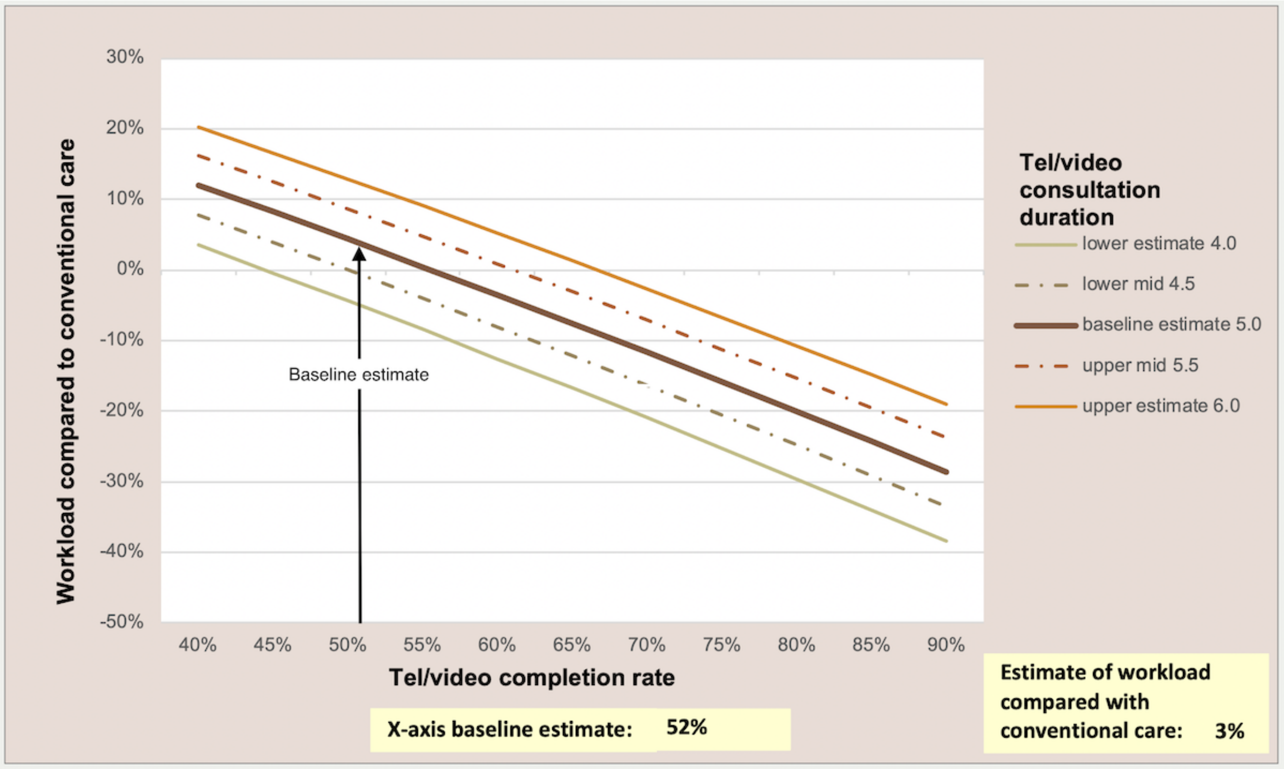
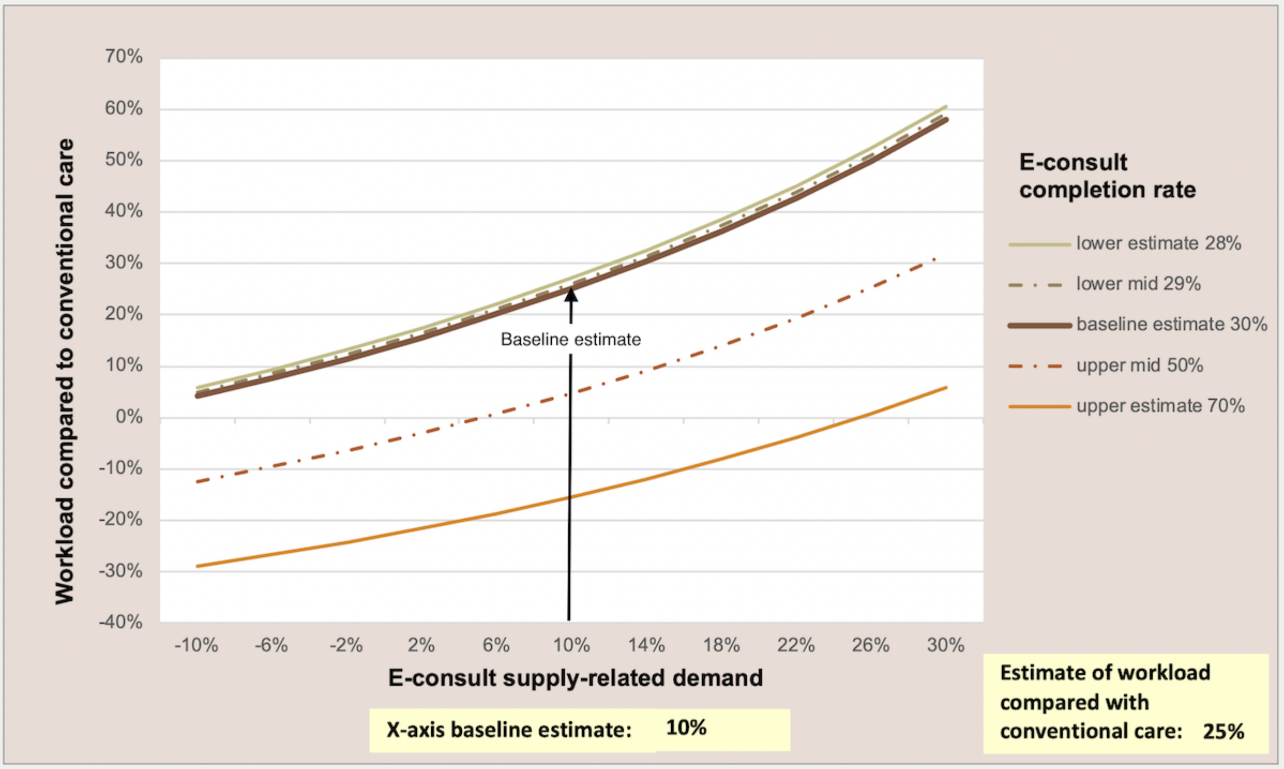


Figure 3. Impact of e-consultations on general practitioner workload: varying supply-related demand and e-consultation completion rate. E-consult: online consultation.



Discussion

Principal Results

Current initiatives to improve access and reduce GP workload through digital-first approaches could have benefits for patients and GPs or could have entirely the opposite effect to that intended. Based on current evidence, these approaches are at least as likely to increase as to decrease workload pressure on general practice. There is potential to reduce general practice workload, but only under stringent conditions, whereby the initial assessment is short and a high proportion of contacts are managed without needing a subsequent face-to-face review. Under almost all scenarios, even modest increases in demand related to improved accessibility would lead to increases in workload.

Comparison With Prior Work

The estimates used to populate this model are based on the best available published evidence. Companies providing digital-first consultation systems claim that general practices using their systems achieve impressive improvements in access and reductions in workload [38,39], and anecdotal reports from some early adopter practices suggest that these benefits can be achieved [4,40]. However, research studies involving larger numbers of practices show that this is not necessarily generalizable [12,29], and UK health services that have introduced digital-first approaches have reported that they do not appear to save GP time [13,41-43]. Furthermore, 2 studies have suggested that telephone-first approaches that lead to reductions in consultations on the same day are compensated for by increased consultations over the next 28 days [11,28]. Evidence from other countries relates to consultation rates or costs rather than workload and provides conflicting findings, with some studies reporting that offering digital consultations leads to increases (including the online consultations themselves) [14,16,26,30,31], while other studies report reductions [18,22,37]. This implies that it is important to understand how and why digital-first approaches are successful in some circumstances but not others. This study helps to inform this debate.

Strengths and Limitations

To ensure comparability between different access pathways, the denominator for our model was patients who decide to seek a consultation with a GP, rather than all patients who use an automated symptom-checker or triage app. About one-third of people look for information online about their symptoms in a conventional consultation system [7], so the use of symptom-checker apps may be partly substituting for this use rather than replacing consultations. The possibility that the availability of an online symptom-checker or triage app could decrease or increase the number of requests for contact with a GP is taken into account using the supply-related demand parameter in our model. Although some providers of online consultation services report that a high proportion of contacts can be managed entirely online [38,39], it is important to consider the denominator, since some of these contacts might not have been made at all or would not have led to a consultation under conventional care. These effects can be considered by comparing the total number of online, telephone, and face-to-face contacts with the number of consultations under a previous conventional approach (see formula in "Systematic Review" in the Methods section).

Some of the estimates used in this model could be challenged because they depend on the context in which the access pathway is introduced. For example, some of the estimates in the model come from the ESTEEM trial of telephone triage of requests for same-day consultations rather than all consultations [11,12]. Second, some clinicians suggest that the duration of a face-to-face consultation might be shorter after a prior online or telephone assessment than under conventional care [19] (although the evidence suggests this is not the case) [23,29]. Third, the proportion of consultations that can be successfully completed online or by telephone might be higher when patients are given the choice to consult in this way, rather than in systems where all patients have to go through this step before accessing a face-to-face consultation. A strength of this study is that by making our model freely available, anyone can explore how workload varies under different assumptions.

The need to limit the scope of the workload model means that it does not consider the effect on other services. The availability of digital consultations could be efficient if it reduces consultations in hospital emergency departments, but the evidence so far provides little support for this hypothesis [12,29,34,44]. The use of digital-first approaches could reduce GP workload by directing patients needing face-to-face care to other primary care professionals, such as nurses and pharmacists. However, we were unable to include this in the model because of a lack of evidence about how much this delegation occurs (compared with the extent to which receptionists direct patients to these professionals under conventional appointment systems) and the proportion of patients that would need to be transferred back to a GP after a nurse or pharmacist consultation. It will be increasingly important to consider the impact on nonmedical workloads in general practice as roles and responsibilities evolve. It is important to note that although delegation to other staff may reduce GP workload, it is not necessarily more efficient for the health care service overall [45].

Finally, we recognize that this study was designed from the perspective of the NHS in the United Kingdom, and usual care is different in other countries. However, by making the workload model freely available, including allowing changes to estimates such as the duration of face-to-face consultations under usual care, we hope that our model will be useful in various settings.

Implications for Clinicians and Policy Makers

If initiatives to improve access to care do lead to increased GP workload, this is not necessarily an argument against them. Ensuring good access to health care is a core purpose of primary care, and additional consultations might represent a response to previously unmet need. However, these initiatives should be justified on the grounds of benefits to patients rather than claims about reductions in GP workload [46]. As with most medical interventions, the key issue is whether the additional benefits are justified in relation to any extra costs. Furthermore, our analysis shows that assumptions about efficiency savings may be misplaced and general practice may need more resources to implement digital-first pathways.

It is important to consider how the benefits of different access pathways are felt by different segments of the population. Digital consultations are predominantly used by patients in the 20-44 age group [17,34], which is a group with generally fewer health care needs. If improving access for them requires more GP time, this will decrease rather than increase the time available for patients with more complex problems, as well as decreasing access for those without internet access. To reduce the potential for worsening health inequalities, it is important to prioritize the use of technology to improve access for the groups of patients with the greatest health care needs, such as older adults, carers, and people with disabilities.

Digital-first access models may have other potential advantages and disadvantages. Improved access could help reverse the decline in public satisfaction with NHS general practice [47] and help avoid inappropriate use of expensive hospital care [3]. Triage systems may offer GPs a greater sense of control over their working day [4]. Technologies to allow GPs to work from home could expand the workforce by unlocking the potential contribution of doctors who cannot work fixed hours in conventional settings [34]. On the other hand, a shift in working patterns toward significant amounts of time spent consulting online or by telephone could lead more GPs to leave the workforce than to join it.

Apart from the impact on workload, there are very important unknowns about the quality and safety of alternatives to face-to-face consultations, as well as the acceptability of these access pathways to different patient groups [2]. These questions should be a priority for research.

Conclusions

This study has highlighted that efficiency gains or losses from the use of digital-first access pathways are finely balanced, and the main impact on workload will be determined by whether these pathways change demand. Digital-first services could increase demand through improved supply or surfacing previously unmet need, or could reduce demand by encouraging patients to self-care or use other services. This is, therefore, an issue of critical importance, but about which we currently have the least evidence. It may take several years for these effects to become manifest. Given that it will be difficult to lower expectations and demand after these have been raised, this suggests the need for careful and staged implementation alongside evaluation rather than universal implementation of digital-first access pathways as soon as possible, as advocated by current UK policy [3].

Acknowledgments

CS is a senior investigator at the National Institute for Health Research. The views expressed in this article are those of the authors, not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

CS had the idea for the paper, reviewed papers, extracted data, developed the first draft of the workload calculator, and wrote the first draft of this paper. MM devised the first draft of the process map for the access pathway and contributed to development of the interactive workload calculator. PD reviewed papers and extracted data for the systematic review. All authors provided critical intellectual input and approved the final paper. CS is guarantor and affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned have been explained.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for literature in MEDLINE (Ovid SP).
[DOCX File , 18 KB - jmir_v22i6e18203_app1.docx]

Multimedia Appendix 2
PRISMA Flow Diagram.
[DOCX File , 38 KB - jmir_v22i6e18203_app2.docx]

Multimedia Appendix 3
Data about model inputs from relevant papers.
[DOCX File , 34 KB - jmir_v22i6e18203_app3.docx]

References

1. NHS England. 2020. Online Consultation Funding URL: https://www.england.nhs.uk/gp/digital-first-primary-care/online-consultation-funding/ [accessed 2020-05-22]
2. Marshall M, Shah R, Stokes-Lampard H. Online consulting in general practice: making the move from disruptive innovation to mainstream service. BMJ 2018 Mar 26;360:k1195. [doi: 10.1136/bmj.k1195] [Medline: 29581174]
3. NHS England. 2019 Jan 31. Investment and evolution: A five-year framework for GP contract reform to implement The NHS Long Term Plan URL: https://www.england.nhs.uk/wp-content/uploads/2019/01/gp-contract-2019.pdf [accessed 2020-05-19]
4. Bakhai M, Croney L, Waller O, Henshall N, Felstead C. NHS England. 2019 Sep 26. Using online consultations in primary care: implementation toolkit URL: https://www.england.nhs.uk/publication/using-online-consultations-in-primary-care-implementation-toolkit/ [accessed 2020-05-22]
5. NHS England. 2016. General Practice Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2016/04/gpfv.pdf [accessed 2020-05-19]
6. Rosen R. Nuffield Trust. 2014 Jun 30. Meeting need or fuelling unnecessary demand? Understanding the impact of improved access to primary care URL: https://www.nuffieldtrust.org.uk/research/meeting-need-or-fuelling-unnecessary-demand-understanding-the-impact-of-improved-access-to-primary-care#partners [accessed 2020-05-19]
7. NHS England. 2019 Jul. GP Patient Survey URL: http://www.gp-patient.co.uk/surveysandreports [accessed 2020-05-19]
8. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. Lancet 2016 Jun 04;387(10035):2323-2330 [FREE Full text] [doi: 10.1016/S0140-6736(16)00620-6] [Medline: 27059888]
9. Salisbury C, Murphy M. University of Bristol. 2020 May. Calculate the impact of online and phone consultations on GP workload URL: http://www.bristol.ac.uk/primaryhealthcare/resources/wlc/ [accessed 2020-05-22]
10. Brunett PH, DiPiero A, Flores C, Choi D, Kum H, Girard DE. Use of a voice and video internet technology as an alternative to in-person urgent care clinic visits. J Telemed Telecare 2015 Jun;21(4):219-226. [doi: 10.1177/1357633X15571649] [Medline: 25697491]
11. Campbell JL, Fletcher E, Britten N, Green C, Holt T, Lattimer V, et al. The clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice: a cluster randomised controlled trial comparing general practitioner-led and nurse-led management systems with usual care (the ESTEEM trial). Health Technol Assess 2015 Feb;19(13):1-212, vii [FREE Full text] [doi: 10.3310/hta19130] [Medline: 25690266]
12. Campbell JL, Fletcher E, Britten N, Green C, Holt TA, Lattimer V, et al. Telephone triage for management of same-day consultation requests in general practice (the ESTEEM trial): a cluster-randomised controlled trial and cost-consequence analysis. Lancet 2014 Nov 22;384(9957):1859-1868 [FREE Full text] [doi: 10.1016/S0140-6736(14)61058-8] [Medline: 25098487]

13. Carter M, Fletcher E, Sansom A, Warren FC, Campbell JL. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. BMJ Open 2018 Feb 15;8(2):e018688 [FREE Full text] [doi: 10.1136/bmjopen-2017-018688] [Medline: 29449293]
14. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009;28(2):323-333. [doi: 10.1377/hlthaff.28.2.323] [Medline: 19275987]
15. de Groot RA, de Haan J, Bosveld HEP, Nijland A, Meyboom-de Jong B. The implementation of a call-back system reduces the doctor's workload, and improves accessibility by telephone in general practice. Fam Pract 2002 Oct;19(5):516-519. [doi: 10.1093/fampra/19.5.516] [Medline: 12356705]
16. Dexter EN, Fields S, Rdesinski RE, Sachdeva B, Yamashita D, Marino M. Patient-Provider Communication: Does Electronic Messaging Reduce Incoming Telephone Calls? J Am Board Fam Med 2016;29(5):613-619 [FREE Full text] [doi: 10.3122/jabfm.2016.05.150371] [Medline: 27613794]
17. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]
18. Ekman B. Cost Analysis of a Digital Health Care Model in Sweden. Pharmacoecon Open 2018 Sep;2(3):347-354 [FREE Full text] [doi: 10.1007/s41669-017-0059-7] [Medline: 29623633]
19. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018 Mar 19;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
20. Gonzalez F, Cimadevila B, Garcia-Comesaña J, Cerqueiro S, Andion E, Prado J, et al. Telephone consultation in primary care. J Health Organ Manag 2018 Apr 09;32(2):321-337. [doi: 10.1108/JHOM-08-2017-0201] [Medline: 29624133]
21. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Sep;69(686):e595-e604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
22. Bergmo TS, Kummervold PE, Gammon D, Dahl LB. Electronic patient-provider communication: will it offset office visits and telephone consultations in primary care? Int J Med Inform 2005 Sep;74(9):705-710. [doi: 10.1016/j.ijmedinf.2005.06.002] [Medline: 16095961]
23. Holt TA, Fletcher E, Warren F, Richards S, Salisbury C, Calitri R, et al. Telephone triage systems in UK general practice: analysis of consultation duration during the index day in a pragmatic randomised controlled trial. Br J Gen Pract 2016 Mar;66(644):e214-e218 [FREE Full text] [doi: 10.3399/bjgp16X684001] [Medline: 26917660]
24. Huygens MWJ, Swinkels ICS, Verheij RA, Friele RD, van Schayck OCP, de Witte LP. Understanding the use of email consultation in primary care using a retrospective observational study with data of Dutch electronic health records. BMJ Open 2018 Jan 21;8(1):e019233 [FREE Full text] [doi: 10.1136/bmjopen-2017-019233] [Medline: 29358442]
25. Jiwa M, Mathers N, Campbell M. The effect of GP telephone triage on numbers seeking same-day appointments. Br J Gen Pract 2002 May;52(478):390-391 [FREE Full text] [Medline: 12014537]
26. Katz SJ, Moyer CA, Cox DT, Stern DT. Effect of a triage-based E-mail system on clinic resource use and patient and physician satisfaction in primary care: a randomized controlled trial. J Gen Intern Med 2003 Sep;18(9):736-744 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.20756.x] [Medline: 12950483]
27. McKinstry B, Hammersley V, Burton C, Pinnock H, Elton R, Dowell J, et al. The quality, safety and content of telephone and face-to-face consultations: a comparative study. Qual Saf Health Care 2010 Aug;19(4):298-303. [doi: 10.1136/qshc.2008.027763] [Medline: 20430933]
28. McKinstry B, Walker J, Campbell C, Heaney D, Wyke S. Telephone consultations to manage requests for same-day appointments: a randomised controlled trial in two practices. Br J Gen Pract 2002 Apr;52(477):306-310 [FREE Full text] [Medline: 11942448]
29. Newbould J, Abel G, Ball S, Corbett J, Elliott M, Exley J, et al. Evaluation of telephone first approach to demand management in English general practice: observational study. BMJ 2017 Sep 27;358:j4197 [FREE Full text] [doi: 10.1136/bmj.j4197] [Medline: 28954741]
30. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health 2014 Mar;20(3):192-198 [FREE Full text] [doi: 10.1089/tmj.2013.0097] [Medline: 24350803]
31. Palen TE, Ross C, Powers JD, Xu S. Association of online patient access to clinicians and medical records with use of clinical services. JAMA 2012 Nov 21;308(19):2012-2019. [doi: 10.1001/jama.2012.14126] [Medline: 23168824]
32. Peleg R. Off-the-cuff cellular phone consultations in a family practice. J R Soc Med 2001 Jun;94(6):290-291 [FREE Full text] [doi: 10.1177/014107680109400610] [Medline: 11387420]
33. Penza KS, Murray MA, Pecina JL, Myers JF, Furst JW. Electronic Visits for Minor Acute Illnesses: Analysis of Patient Demographics, Prescription Rates, and Follow-Up Care Within an Asynchronous Text-Based Online Visit. Telemed J E Health 2018 Mar;24(3):210-215. [doi: 10.1089/tmj.2017.0091] [Medline: 28731388]

34. Quigley A, Hex N, Aznar C, Salisbury C. NHS Hammersmith and Fulham CCG and NHS England. 2019 May. Evaluation of Babylon GP at hand: Final evaluation report URL: https://www.hammersmithfulhamccg.nhs.uk/media/156123/Evaluation-of-Babylon-GP-at-Hand-Final-Report.pdf [accessed 2020-05-19]
35. Richards DA, Meakins J, Tawfik J, Godfrey L, Dutton E, Richardson G, et al. Nurse telephone triage for same day appointments in general practice: multiple interrupted time series trial of effect on workload and costs. BMJ 2002 Nov 23;325(7374):1214 [FREE Full text] [doi: 10.1136/bmj.325.7374.1214] [Medline: 12446539]
36. Stevens S, Bankhead C, Mukhtar T, Perera-Salazar R, Holt TA, Salisbury C, NIHR School for Primary Care Research. Patient-level and practice-level factors associated with consultation duration: a cross-sectional analysis of over one million consultations in English primary care. BMJ Open 2017 Nov 16;7(11):e018261 [FREE Full text] [doi: 10.1136/bmjopen-2017-018261] [Medline: 29150473]
37. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
38. Longman H. askmyGP. 2017 Oct 08. Cut your workload by 10%, without fail URL: https://askmygp.uk/cut-workload-10-without-fail/ [accessed 2020-05-19]
39. eConsult is the most used online consultation platform by NHS GP practices in the UK. eConsult URL: https://econsult.net/primary-care/ [accessed 2020-05-19]
40. NHS England. Releasing time for care URL: https://www.england.nhs.uk/gp/gpfv/redesign/gpdp/releasing-time/ [accessed 2020-05-19]
41. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2018 Jan;68(666):e1-e8 [FREE Full text] [doi: 10.3399/bjgp17X693509] [Medline: 29109115]
42. Patel A. Pulse. 2019 Aug 21. No evidence that online consultations save time, says CCG URL: http://www.pulsetoday.co.uk/news/all-news/no-evidence-that-online-consultations-save-time-says-ccg/20039267.article [accessed 2020-05-19]
43. Lawless M, Wright E, Davidson J. A collaborative approach to improving patient access in general practice: impact of three different pilot schemes in 12 general practices in Greenwich. London J Prim Care (Abingdon) 2016;8(4):56-65 [FREE Full text] [doi: 10.1080/17571472.2016.1173946] [Medline: 28250835]
44. Bunn F, Byrne G, Kendall S. The effects of telephone consultation and triage on healthcare use and patient satisfaction: a systematic review. Br J Gen Pract 2005 Dec;55(521):956-961 [FREE Full text] [Medline: 16378566]
45. Hollinghurst S, Horrocks S, Anderson E, Salisbury C. Comparing the cost of nurse practitioners and GPs in primary care: modelling economic data from randomised trials. Br J Gen Pract 2006 Jul;56(528):530-535 [FREE Full text] [Medline: 16834880]
46. Downey A. Digital Health. Digital solutions 'should be encouraged' to ease pressure on GPs URL: https://www.digitalhealth.net/2019/08/digital-solutions-ease-pressure-gps/ [accessed 2020-05-19]
47. Robertson R, Appleby J, Evans H. Nuffield Trust. 2018 Feb. Public satisfaction with the NHS and social care in 2017: Results and trends from the British Social Attitudes survey URL: https://www.nuffieldtrust.org.uk/files/2018-02/nut-kf-bsa-2018-web.pdf [accessed 2020-05-19]

Abbreviations

GP: general practitioner

Edited by G Eysenbach; submitted 11.02.20; peer-reviewed by M Carter, F North; comments to author 20.03.20; revised version received 26.03.20; accepted 28.03.20; published 16.06.20.

Please cite as:
Salisbury C, Murphy M, Duncan P
The Impact of Digital-First Consultations on Workload in General Practice: Modeling Study
J Med Internet Res 2020;22(6):e18203
URL: http://www.jmir.org/2020/6/e18203/
doi:10.2196/18203
PMID:32543441

Original Paper

Applying Machine Learning to Daily-Life Data From the TrackYourTinnitus Mobile Health Crowdsensing Platform to Predict the Mobile Operating System Used With High Accuracy: Longitudinal Observational Study

Rüdiger Pryss[1], PhD, Prof Dr; Winfried Schlee[2], PhD; Burkhard Hoppenstedt[3], MSc; Manfred Reichert[3], PhD, Prof Dr; Myra Spiliopoulou[4], PhD, Prof Dr; Berthold Langguth[2], PhD, Prof Dr; Marius Breitmayer[3], MSc; Thomas Probst[5], PhD, Prof Dr

[1]Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany
[2]Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany
[3]Institute of Databases and Information Systems, Ulm University, Ulm, Germany
[4]Faculty of Computer Science, Otto von Guericke University of Magdeburg, Magdeburg, Germany
[5]Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, Austria

Corresponding Author:
Rüdiger Pryss, PhD, Prof Dr
Institute of Clinical Epidemiology and Biometry
University of Würzburg
Josef-Schneider-Str 2
Würzburg, 97080
Germany
Phone: 49 931 20146471
Email: ruediger.pryss@uni-wuerzburg.de

Abstract

Background: Tinnitus is often described as the phantom perception of a sound and is experienced by 5.1% to 42.7% of the population worldwide, at least once during their lifetime. The symptoms often reduce the patient's quality of life. The TrackYourTinnitus (TYT) mobile health (mHealth) crowdsensing platform was developed for two operating systems (OS)—Android and iOS—to help patients demystify the daily moment-to-moment variations of their tinnitus symptoms. In all platforms developed for more than one OS, it is important to investigate whether the crowdsensed data predicts the OS that was used in order to understand the degree to which the OS is a confounder that is necessary to consider.

Objective: In this study, we explored whether the mobile OS—Android and iOS—used during user assessments can be predicted by the dynamic daily-life TYT data.

Methods: TYT mainly applies the paradigms ecological momentary assessment (EMA) and mobile crowdsensing to collect dynamic EMA (EMA-D) daily-life data. The dynamic daily-life TYT data that were analyzed included eight questions as part of the EMA-D questionnaire. In this study, 518 TYT users were analyzed, who each completed at least 11 EMA-D questionnaires. Out of these, 221 were iOS users and 297 were Android users. The iOS users completed, in total, 14,708 EMA-D questionnaires; the number of EMA-D questionnaires completed by the Android users was randomly reduced to the same number to properly address the research question of the study. Machine learning methods—a feedforward neural network, a decision tree, a random forest classifier, and a support vector machine—were applied to address the research question.

Results: Machine learning was able to predict the mobile OS used with an accuracy up to 78.94% based on the provided EMA-D questionnaires on the assessment level. In this context, the daily measurements regarding how users concentrate on the actual activity were particularly suitable for the prediction of the mobile OS used.

Conclusions: In the work at hand, two particular aspects have been revealed. First, machine learning can contribute to EMA-D data in the medical context. Second, based on the EMA-D data of TYT, we found that the accuracy in predicting the mobile OS used has several implications. Particularly, in clinical studies using mobile devices, the OS should be assessed as a covariate, as it might be a confounder.

(J Med Internet Res 2020;22(6):e15547) doi:10.2196/15547

KEYWORDS

mHealth; crowdsensing; tinnitus; machine learning; mobile operating system differences; ecological momentary assessment; mobile phone

Introduction

Background

Mobile health (mHealth) uses smart mobile devices to address various questions in the context of neuroscience, psychology, and medicine. New paradigms, such as ecological momentary assessment (EMA), mobile crowdsourcing, and mobile crowdsensing, as well as mHealth apps, in general, have enabled data collection procedures that surpass many existing methods in gathering valuable medical data by several orders of magnitude [1]. Among others, by using smart mobile devices, data can be gathered in everyday life, on a cost-effective basis, and by adding contextual information sources, such as Twitter or Facebook. As many medical phenomena pose daily variations [2], mHealth technology is predestined to be utilized in this context. Along these trends, many insights have been presented by researchers that show that smart mobile devices can help to establish new data sources in many scenarios [3].

In these data collection scenarios, which are built on the usage of mobile devices and their sensors, one dimension has been less considered so far. It refers to the question of whether the operating system (OS) of the mobile technology being used (eg, iOS or Android) constitutes a valuable information source or confounder for medical data analyses. Or, as another example, is it possible to derive insights if a patient changes the OS during a study when using mHealth apps? As Android and iOS dominate the mobile OS market [4]—with a market share of 99.32% in May 2020 (72.52% Android and 26.80% iOS)—any insights gained based on differences from users regarding these OS types could provide a representative picture for the OS market. Following this, data that were gathered with the TrackYourTinnitus (TYT) mHealth crowdsensing platform for tinnitus patients over 5 years of age are analyzed in this paper. TYT is an mHealth crowdsensing platform that offers iOS and Android apps that can empower patients to learn more about their tinnitus symptoms over time. Tinnitus is the phantom perception of a sound and it is experienced by 5.1% to 42.7% of the population worldwide at least once during their lifetime [5]. The symptoms often reduce the patient's quality of life. As tinnitus constitutes a chronic condition for which currently no cure or general treatment exists, patients suffering from it crave for new treatment procedures or at least new medical insights. With the idea of EMA, also known as ambulatory assessment or experience sampling, and mobile crowdsensing techniques in mind, TYT was developed by an interdisciplinary team of medical experts, psychologists, and computer scientists.

The development of TYT was motivated by the clinical experience that among many tinnitus patients, tinnitus loudness and tinnitus annoyance vary over time and that patients' experiences differ in the pattern of these fluctuations. Therefore, the variations are considered to provide new valuable insights in the pathophysiological mechanisms of this chronic condition [6]. To learn more about these fluctuations, TYT applies EMA and mobile crowdsensing to capture them. In EMA, the variable in question (eg, a symptom) is assessed repeatedly in daily life [7]. In mobile crowdsensing, only mobile devices are used for the data collection procedure, while their sensors are used to capture, for example, the GPS position or the external sound level [8]. In contrast, in mobile crowdsourcing, tasks are proposed by a crowdsourcer to a group of individuals, who voluntarily undertake tasks. The undertaking of the task always entails mutual benefit. The user will receive the satisfaction of a given type of need, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture [9]. In contrast to mobile crowdsourcing, mobile crowdsensing relies solely on mobile technology and integrates sensors to collect data. Two recent works that discuss mobile crowdsensing in the context of health care can be found in Kraft et al [1] and Pryss [10]. In TYT, the users fill in a registration questionnaire (ie, static data) and can provide repeated assessments in daily life (ie, dynamic data) afterward [11].

Objectives

Compared to the existing studies on TYT, this work investigates repeatedly provided EMA datasets from TYT users (ie, dynamic data) and their relation to the mobile OS used. While this study analyzes this dynamic data, a previous study focused on differences between Android and iOS users in the static data given at registration [12]. Contrary to the Android versus iOS comparison of the SmokeFree28 (SF28) smoking cessation app [13], in our study we found no differences in gender, but we did find differences in age for TYT users. However, in Pryss et al [12], we found differences that might be of interest for medical purposes. More specifically, we revealed that Android users reported a significantly longer tinnitus duration than did iOS users, cross-sectionally. Future longitudinal research is necessary to address the question of whether users with longer tinnitus duration prefer Android to iOS or whether users of Android tend to develop longer tinnitus durations than iOS users. In another recent work [14], we investigated differences in Android and iOS users of the TrackYourHearing (TYH) mHealth crowdsensing platform. This platform aims to measure fluctuations in hearing of users with hearing loss. In the TYH study, we found no differences in gender or age, but significant differences were revealed in three questions of the dynamic data that were repeatedly provided. This shows that the dynamic data in combination with the OS are worth being investigated more deeply.

As another current trend, the application of machine learning techniques in different fields is promising. In the medical field, there is a remarkable discrepancy between huge expectations in the potential of machine learning on one side and the current application of this technique on the other [15-19]. Importantly, there is an increasing consensus about its potential in the context of mobile technology [20-23]. However, the application of machine learning to a large group of users of an mHealth crowdsensing platform that gathers EMA datasets is still rare [19,24-27]. As we already found relevant differences between Android and iOS pertaining to the TYT users' static characteristics at registration [12], this work investigates the following research question: *Is it possible to predict the mobile OS used based on dynamic TYT data with high accuracy using machine learning methods? More specifically, is it possible to predict the mobile OS used based on the repeatedly given daily data provided by the TYT users with high accuracy using machine learning methods?* To the best of our knowledge, thus far, no other work has considered this research question in the given context.

Methods

Overview

TYT was developed to track the individual tinnitus perception of users in their daily lives [28]. In this context, the procedure shown in Figure 1 is applied to all TYT users. In general, TYT pursues three major goals.

First, dynamic EMA (EMA-D) data shall be collected during the continuous mobile crowdsensing procedure (see Figure 1, box #4). Importantly, a crowdsensing user shall not foresee the times he or she is asked to provide the data (see Figure 1, box #3). This is ensured by asking the crowdsensing users for data in various daily-life situations by the use of smartphone notifications. When a user clicks on such a notification, the tinnitus-tracking questionnaire is presented to a user, consisting of eight EMA-D questions. Table 1 lists the eight questions of the EMA-D questionnaire.

Note that the questionnaire appears visually on both mobile OS types in the same way. For more information on the questionnaire shown in Table 1 and how it appears on the mobile devices, see Pryss et al [11].

Second, the collected data shall enable innovative data analyses, such as juxtaposing the prospectively assessed EMA-D and retrospectively assessed static EMA (EMA-S) at registration (see Figure 1, box #2; [11]). Third, gathered data shall be used to provide feedback to the mobile crowdsensing users [29].

When initially designing the user procedure of TYT, we had not yet considered comparing users based on the mobile OS they used. The initial intention to collect information about the mobile OS used (see Figure 1, box #1) when filling out a questionnaire had been to quickly identify technical issues that could emerge with the large variety of mobile OS versions and mobile devices used. However, it turned out that the information can be also used for innovative analyses. For interested readers, more technical information of the platform can be found in Pryss et al [28,29].

A further note is provided to distinguish between static and dynamic data in the procedure shown in Figure 1. Usually, existing works distinguish between questionnaire, sensor, and behavioral data when utilizing mHealth crowdsensing approaches [30-32]. However, our distinction between static *trait* (ie, EMA-S) or dynamic *state* EMA data (ie, EMA-D) is done less frequently by other works. This is remarkable, as the distinction between trait (ie, static) and state (ie, dynamic) variables is fundamental in clinical and psychological research. As an example, trait data are expected to have a closer association with genetic information as compared to state data, which depend more strongly on environmental factors.

The experimental protocols were approved by the Ethics Committee of the University Clinic of Regensburg, Germany. All methods were carried out in accordance with the relevant guidelines and regulations. The users of the app were informed that their gathered data will be used for scientific analyses; informed consent was given.

Figure 1. TrackYourTinnitus (TYT) mobile crowdsensing collection procedure. EMA: ecological momentary assessment.

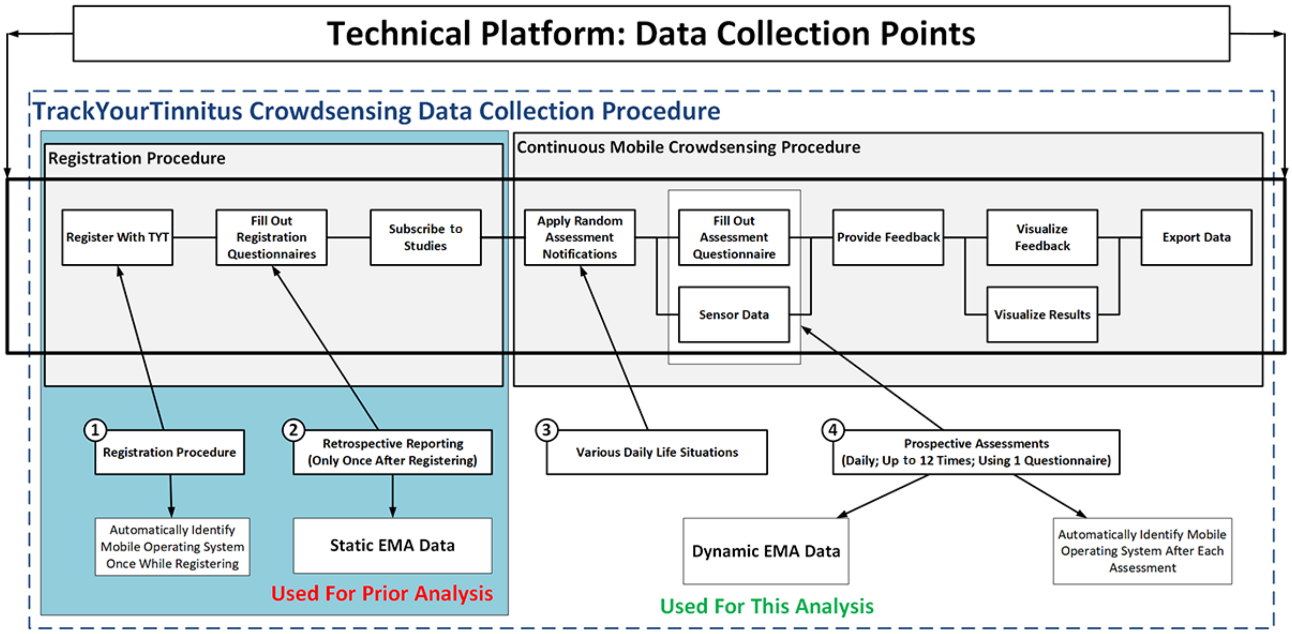


Table 1. Questions from the dynamic ecological momentary assessment (EMA-D) questionnaire.

Number	Question (Q)	Answer type
Q1	Did you perceive the tinnitus right now?	Yes or no
Q2	How loud is the tinnitus right now?	Slider[a]
Q3	How stressful is the tinnitus right now?	Slider
Q4	How is your mood right now?	Manikins[b]
Q5	How is your arousal right now?	Manikins
Q6	Do you feel stressed right now?	Slider
Q7	How much did you concentrate on the things you are doing right now?	Slider
Q8	Do you feel irritable right now?	Yes or no

[a] Each slider has a different range; the slider for Q2, for example, ranges from *not audible* to *maximal loudness*.

[b] We made use of the Self-Assessment Manikin (SAM) scales, which are a pictorial rating system to obtain self-assessments of experienced emotions on the dimensions affective valence, dominance, and arousal.

Data Source

The TYT platform includes a website, uses a relational database, and includes an iOS and Android app. The latter are implemented as native apps. Users can register with the platform by using the website or the mobile apps. After that, three registration questionnaires must be completed—EMA-S questionnaires, which can be filled out using the website or the mobile apps—before users can provide the EMA-D data repeatedly in daily life—this is denoted as the EMA-D questionnaire, which can only be filled out using the mobile apps. After completing the registration questionnaires, users must decide whether they want to use the default notification schema for the EMA-D questionnaire. The default setting means users would receive random notifications up to eight times per day. This setting can be changed by a user in many ways. The user can reduce the notification number to a minimum value of three or a maximum value of 12 notifications per day. In addition, a user can select specific days of the week when no notifications shall appear. Finally, a user can switch to the fixed notification mode, in which he or she specifies exact notification points. Note that in this analysis, it is not distinguished which mode has been selected by a user. Finally, if the user clicks on a notification, the EMA-D questionnaire appears. A detailed description can be found in Pryss et al [11]. Finally, note that users can fill out the EMA-D questionnaire in a user-initiated manner as well (ie, without getting a notification to fill out a questionnaire).

Another feature is offered to the TYT users. They can obtain their results of all answered EMA-D questionnaires through the apps or the website. For this purpose, two options are provided: first, they can visualize the results via the website or the mobile apps; or second, they can download a CSV (comma-separated values) file, only via the website, for further personal evaluations.

Participants

The analysis was conducted in March 2020. At this time, the TYT platform had 4835 registered users. From them, 2584 users completed the EMA-D questionnaire at least once and, in total, 75,278 EMA-D questionnaires were available. To get an impression of how TYT is used worldwide, the country distribution was determined; it shows the number of completed EMA-D questionnaires (ie, all eight items filled in) from 2065 users from the 12 countries with the most completed EMA-D questionnaires out of the 2584 users who completed the questionnaires. This resulted in 67,789 EMA-D questionnaires from 2065 users. The worldwide distribution is shown in Table 2.

The OS distribution of the 2584 users who completed the EMA-D questionnaire at least once is as follows: 40.02% (1034/2584) of the data were provided by iOS users, while 59.98% (1550/2584) were provided by Android users. The OS distribution of all completed EMA-D questionnaires in TYT is as follows: 32.00% (24,089/75,278) of the data were provided by iOS users, while 68.00% (51,189/75,278) were provided by Android users.

The data preparation steps for the machine learning analysis, including use of a scikit-learn function [33] to compare the same number of EMA-D questionnaires from Android and iOS users, can be seen in Figure 2.

For the final study sample of 297 Android users and 221 iOS users, Table 3 shows statistical comparisons between the Android and iOS users in terms of gender, age, and numbers of completed EMA-D questionnaires (chi-square test and *t* tests for independent samples, two-sided). Age was set to *missing* if users provided invalid entries.

Finally, Figure 3 shows the histogram for the number of completed EMA-D questionnaires for the 518 investigated TYT users (see Figure 2).

Table 2. Country distribution of TrackYourTinnitus (TYT) users (n=2065) in ascending order.

Number	Country	Completed dynamic ecological momentary assessment (EMA-D) questionnaires (n=67,789), n (%)
1	Australia	535 (0.79)
2	Belgium	819 (1.21)
3	Italy	1026 (1.51)
4	Russia	1076 (1.59)
5	Austria	1110 (1.64)
6	Norway	1159 (1.71)
7	Canada	2113 (3.12)
8	Great Britain	3202 (4.72)
9	Switzerland	5229 (7.71)
10	Netherlands	6917 (10.20)
11	United States	9117 (13.45)
12	Germany	35,486 (52.35)

Figure 2. Data preparation steps for the machine learning analysis. [a]Information about the scikit-learn function can be found on the scikit-learn website [33]. EMA-D: dynamic ecological momentary assessment; Q: question; TYT: TrackYourTinnitus.

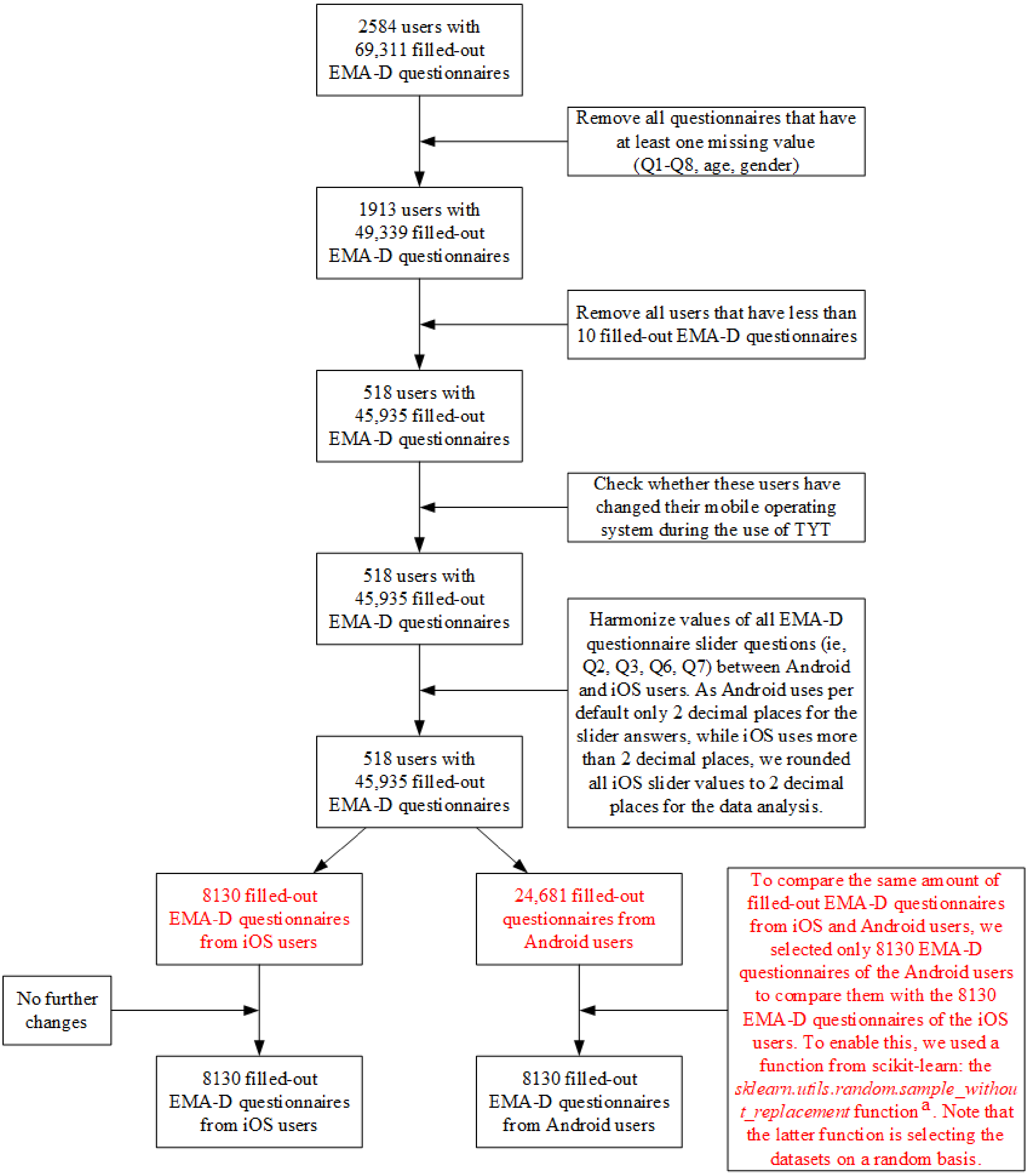


Table 3. Comparisons between iOS and Android users regarding gender, age, and number of completed dynamic ecological momentary assessment (EMA-D) questionnaires.

Variable	Android	iOS	Chi-square (*df*)	Two-tailed *t* test (*df*)	*P* value
Gender[a] (male), n (%)	221 (74.4)	147 (66.5)	1.2 (1)	N/A[b]	.27
Age[c] (years), mean (SD)	49.76 (12.29)	50.57 (13.09)	N/A	–0.71 (497)	.48
Number of completed EMA-D questionnaires[d] (ie, all eight questions completed), mean (SD)	49.52 (80.32)	66.55 (125.92)	N/A	–1.87 (516)	.06

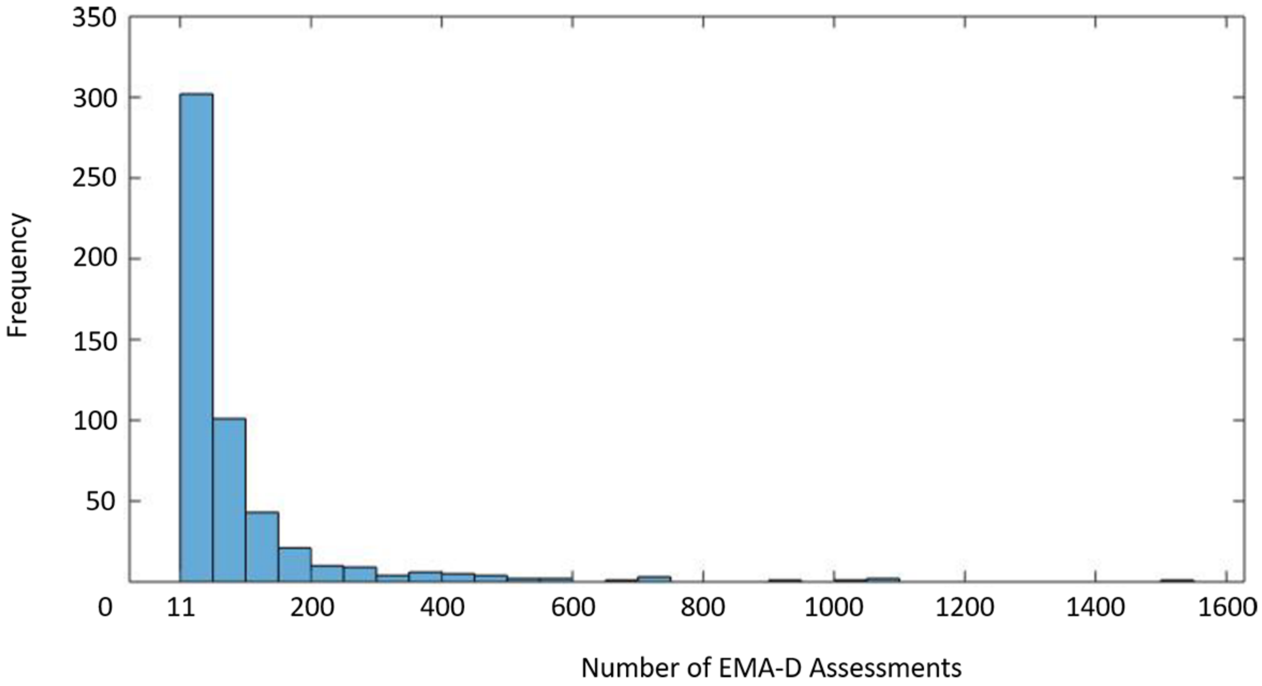
[a] Sample sizes for gender are n=297 for Android and n=221 for iOS.

[b] N/A: not applicable.

[c] Sample sizes for age are n=295 for Android and n=204 for iOS.

[d] Sample sizes for number of completed EMA-D questionnaires are n=297 for Android and n=221 for iOS.

Figure 3. Frequencies of completed questionnaires of the investigated TrackYourTinnitus (TYT) users (n=518). EMA-D: dynamic ecological momentary assessment.



Machine Learning Analysis

We applied machine learning approaches with the goal to predict the OS—Android or iOS—of a provided assessment of the EMA-D data. For this purpose, four machine learning approaches were applied to the dataset: a feedforward neural network (FNN), a decision tree (DT), a support vector machine (SVM), and a random forest classifier (RFC). All approaches were chosen because they are appropriate for high-dimensional datasets, which is the case for the given EMA-D questionnaires of the TYT users. This is supported by similar works [34,35].

Concerning the dataset in general, it is noteworthy that the machine learning approaches were applied on the assessment level of the EMA-D questionnaires. This means that assessments from one user can be in both the training and the validation datasets. Performing a separation on assessment level has advantages and disadvantages. As the main disadvantage, it can be argued that if a participant is in both datasets, then there might be a bias. On the other hand, if users of the training phase are separated from the validation phase users, then it must be ensured that the user characteristics between the training and validation phases generate no bias. In an EMA-driven approach, where daily assessments on a random and voluntary basis are the main goal, it is difficult to be able to evaluate a large group of users with similar assessment characteristics. However, in future work, it will be a further goal to also separate the dataset on the user level in a reasonable manner. That individual users play an important role in health care studies is emphasized by the emerging paradigm of N-of-1 studies [36].

Concerning the validation procedure, the following has been accomplished: in a first step, the validation was based on a *10-fold cross-validation* approach [34,35] (ie, for the SVM, the DT, and the RFC). Here, the entire dataset was distributed into 10 equal parts. Nine of these parts were used for the training phase, while the remaining one part was used for the testing phase. The whole procedure was repeated 10 times and the average values were then calculated over all 10 runs. To foster our results, another validation was performed for the SVM, the

DT, and the FNN. We conducted a leave-one-out approach on the user level, for all of the 518 included users, combined with a majority vote for the EMA-D questionnaires from the user that was left out, to see whether the prediction differs if the EMA-D questionnaires from the user who was left out are excluded. In conclusion, there was no obvious difference observed.

For all analyses, the following technical environment was used: a laptop with an i7 core (2.60 GHz); MATLAB, version R2017a (MathWorks); the Statistics and Machine Learning Toolbox (MathWorks); and scikit-learn, open source machine learning library, for Python (Python Software Foundation). For all applied methods, we used the default parameters of the technical environment. In MATLAB, the FNN, the SVM, and the DT were calculated, while in Python scikit-learn, the RFC was calculated.

Results

Overview

The four applied machine learning approaches showed different results for the research question. In general, the prediction accuracies were unexpectedly high.

Feedforward Neural Network

The FNN was the worst-performing candidate. Here, for 72.67% of the EMA-D questionnaires, the mobile OS could be correctly predicted. In the MATLAB toolbox that was used, the essential parameter for the calculation was *feedforwardnet(10)*.

Decision Tree

The DT, in turn, performed as the third-best candidate. It was applied with a 10-fold cross-validation and it predicted the correct mobile OS for 76.36% of the EMA-D questionnaires. Importantly, the resulting DT has a depth of 379, showing that the prediction can be categorized into a high-dimensional calculation. In the MATLAB toolbox that was used, the essential parameter for the calculation was *fitctree(X,Y,'CrossVal','on')*.

Random Forest Classifier

The RFC performed as the best candidate; the mobile OS could be predicted correctly for 78.94% of the EMA-D questionnaires. In the Python scikit-learn method that was used, the essential parameter for the calculation was as follows: *RandomForestClassifier(n_estimators=100, bootstrap=True, max_features='sqrt', random_state=42)*. In addition to the prediction results, Table 4 shows the importance of the eight EMA-D questions for the overall prediction result of 78.94%; here, we used the *model.feature_importances_* feature of Python scikit-learn. Importantly, question 7 and then question 2 are the most important questions for the prediction result of 78.94%.

Table 4. Importance of the eight dynamic ecological momentary assessment (EMA-D) questions for the random forest classifier prediction. Question 1 (Q1): Did you perceive the tinnitus right now? (yes or no); Q2: How loud is the tinnitus right now? (slider); Q3: How stressful is the tinnitus right now? (slider); Q4: How is your mood right now? (manikins); Q5: How is your arousal right now? (manikins); Q6: Do you feel stressed right now? (slider); Q7: How much did you concentrate on the things you are doing right now? (slider); and Q8: Do you feel irritable right now? (yes or no).

Question Number	Q1	Q8	Q5	Q4	Q3	Q6	Q2	Q7
Percentage of Importance	0.03043	0.03985	0.08728	0.0913	0.17246	0.17425	0.19247	0.21194

Support Vector Machine

The SVM performed as the second-best candidate. Overall, using all eight questions, the mobile OS could be predicted correctly for 78.65% of the EMA-D questionnaires. For the SVM, detailed results for single questions and question combinations are discussed in more detail. This will show that all eight questions are needed to get a prediction result with an accuracy that shows that the OS might be a confounder that should be further considered. The same detailed discussion could be accomplished for the other approaches, such as the RFC. We opted for the SVM for a more detailed discussion and to compare the results to other approaches to see if they deviate significantly from each other. More specifically, prediction results for combinations of two questions as well as single questions are shown in Table 5. Seven results will be further discussed. The discussion will show that the accuracies vary among the eight EMA-D questions on one hand. One the other hand, it will show that despite the observed variances, the overall achieved accuracy is high for different questions and their combinations.

First, we consider question 7—How much did you concentrate on the things you are doing right now? (slider)—and question 8—Do you feel irritable right now? (yes or no). They performed as the two best single questions for the prediction. Each of them has an accuracy of 58.80%. This result is only partly confirmed by the RFC. For the RFC, question 7 is also very important, but question 8 is less important for the RFC.

Second, question 5—How is your arousal right now? (manikins)—performed with the third-best result for a single question; here, an accuracy of 57.14% was attained. This question is like question 8, in that it is less important in the case of the RFC.

Third, the combination of question 7—How much did you concentrate on the things you are doing right now? (slider)—and question 8—Do you feel irritable right now? (yes or no)—performed as the best candidate for two-question combinations; in this case, an accuracy of 63.95% was achieved. This result is again only partly supported by the RFC (ie, for the RFC, question 8 was less important; see Table 4).

Fourth, the worst result was achieved when only using question 4—How is your mood right now? (manikins)—as the predictor. For question 4, an accuracy of 54.07% was achieved. Again, this deviates from the result of the RFC, where question 1 was the worst candidate.

Fifth, when solely combining *yes or no* questions (ie, question 1—Did you perceive the tinnitus right now?—and question 8—Do you feel irritable right now?), the mobile OS could be predicted correctly for 63.37% of the user assessments. This result also shows that without slider questions, a meaningful accuracy can be achieved.

Sixth, when looking at question-question combinations that include only sliders as answer types, the highest accuracies were achieved by the combination of question 2—How loud is the tinnitus right now? (slider)—and question 7—How much did you concentrate on the things you are doing right now? (slider). Here, an accuracy of 59.86% was achieved. This, in turn, is supported by the result of the RFC.

Seventh, it is remarkable that the overall prediction result with all eight questions is considerably higher than with single questions or combinations of two questions.

Finally, Table 6 represents the confusion table for the SVM calculations. Note that the values are for all eight EMA-D questions of the considered 14,708 Android questionnaires as well as 14,708 iOS EMA-D questionnaires.

Table 5. Prediction accuracies of the support vector machine (SVM) based on the eight dynamic ecological momentary assessment (EMA-D) questions and their combinations.

Question[a] (Q)	Accuracy for each question (Q) combination, %							
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Q1	55.69	—	—	—	—	—	—	—
Q2	59.18	55.90	—	—	—	—	—	—
Q3	58.53	56.80	56.61	—	—	—	—	—
Q4	56.37	57.89	58.28	54.07	—	—	—	—
Q5	59.55	61.08	61.31	60.10	57.14	—	—	—
Q6	58.59	57.27	56.35	58.69	62.83	56.28	—	—
Q7	61.24	59.86	59.40	60.19	62.38	59.33	58.80	—
Q8	63.37	62.32	63.57	60.18	61.57	62.67	63.95	58.80

[a] Q1: Did you perceive the tinnitus right now? (yes or no); Q2: How loud is the tinnitus right now? (slider); Q3: How stressful is the tinnitus right now? (slider); Q4: How is your mood right now? (manikins); Q5: How is your arousal right now? (manikins); Q6: Do you feel stressed right now? (slider); Q7: How much did you concentrate on the things you are doing right now? (slider); and Q8: Do you feel irritable right now? (yes or no).

Table 6. Confusion table for the support vector machine (SVM) calculations over all eight dynamic ecological momentary assessment (EMA-D) questions.

Predicted class	Actual class			
	iOS		Android	
	True positives, n	False negatives, n	False positives, n	True negatives, n
iOS	13,002	N/A[a]	1967	N/A
Android	N/A	1706	N/A	12,741

[a] N/A: not applicable.

Importance of Questions

In general, the question emerges as to why some of the eight EMA-D questions are better suited than others to correctly predict the mobile OS. One possible explanation refers to the answering behavior of the users of the two mobile OS types. To illustrate this, Figures 4 and 5 show, as examples, histograms of question 3—How stressful is the tinnitus right now? (slider)—and question 5—How is your arousal right now? (manikins). It is obvious that Android and iOS users answer differently.

Figure 4. Answers to question 3—How stressful is the tinnitus right now? (slider)—and the difference vector of Android and iOS users.

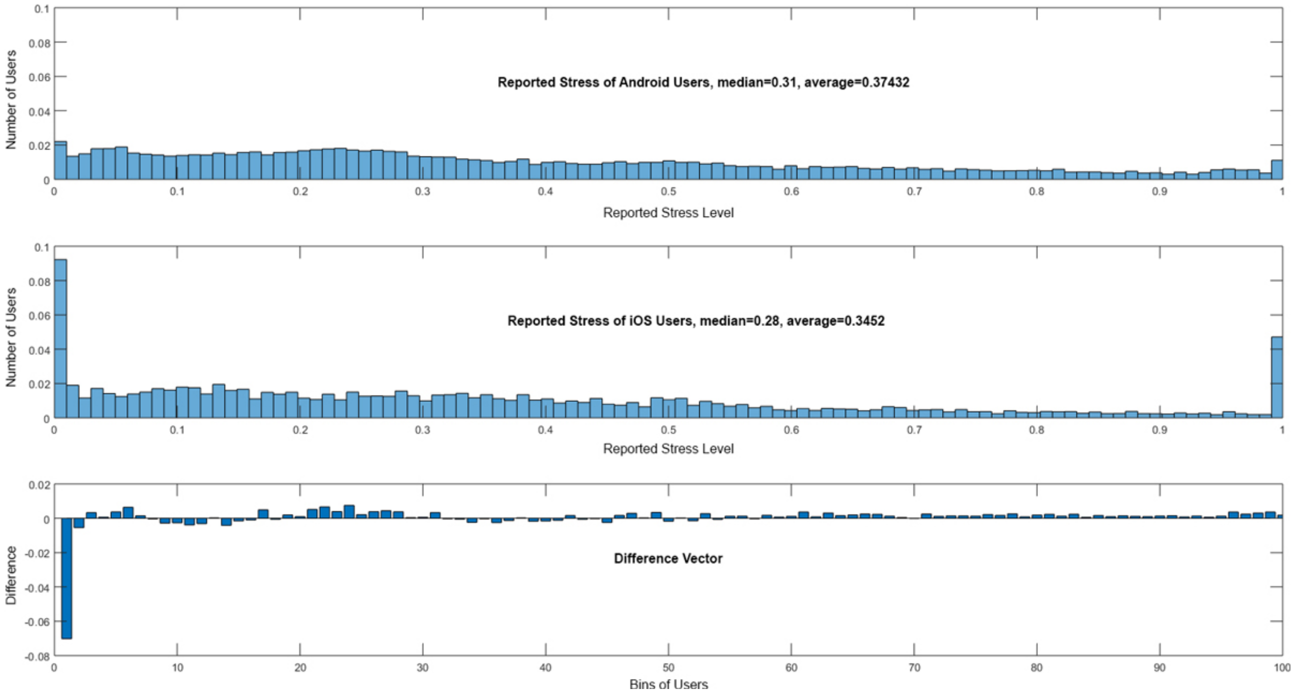
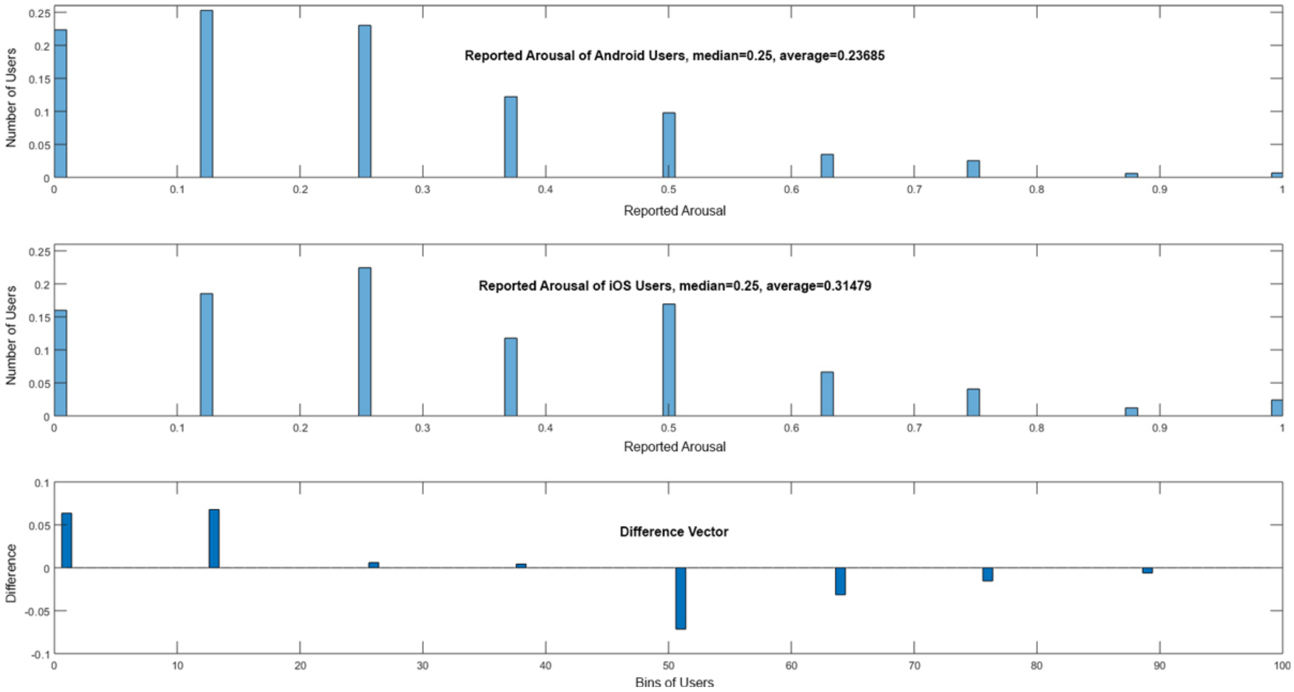


Figure 5. Answers to question 5 "How is your arousal right now? (manikins)" and the difference vector of Android and iOS users.



Question 3 and question 5 have been chosen as examples, as they differ in their importance between the machine learning approaches: question 3 is the fourth-most important in SVM versus the third-most important in RFC, while question 5 is the third-most important in SVM versus the sixth-most important in RFC. In addition, other questions are more suitable for the overall prediction. Although they differ and other questions are better, they still show striking differences between assessments from Android and iOS users. To support this result, further consider Figures 6 and 7; they each show data for 100 users in total, distributed among Android and iOS. The data were randomly selected out of the entire dataset. This subset was chosen for the sake of clarity; if all data points were shown, less could be visually observed. In Figure 6, for question 2—How loud is the tinnitus right now? (slider)—in combination with question 3—How stressful is the tinnitus right now? (slider)—shown on the left-hand side of the figure, or question 6—Do you feel stressed right now? (slider)—shown on the right-hand side of the figure, the blue dots show the answers from the Android users, while the red dots show answers from the iOS users. It is striking that Android and iOS users answer differently. Furthermore, in Figure 7, for question 4—How is your mood right now? (manikins)—in combination with question 3—How stressful is the tinnitus right now?

(slider)—shown on the left-hand side of the figure, or question 5—How is your arousal right now? (manikins)—shown on the right-hand side of the figure, the same can be observed. Importantly, Figures 6 and 7 are not representative of the entire dataset, but it is nevertheless notable that Android and iOS users answer differently. Further note that in Figures 6 and 7, we do not illustrate the achieved predictions. Instead, the attained loss is shown (ie, *1-loss* denotes the achieved accuracy). Furthermore, these combinations have been selected as they also show clear differences between Android and iOS assessments, although other questions have higher prediction accuracies.

Figure 6. Support vector machine (SVM) results for question 2 (Q2), combined with question 3 (Q3) and question 6 (Q6), from 100 data entries. Q2: How loud is the tinnitus right now? (slider); Q3: How stressful is the tinnitus right now? (slider); and Q6: Do you feel stressed right now? (slider).

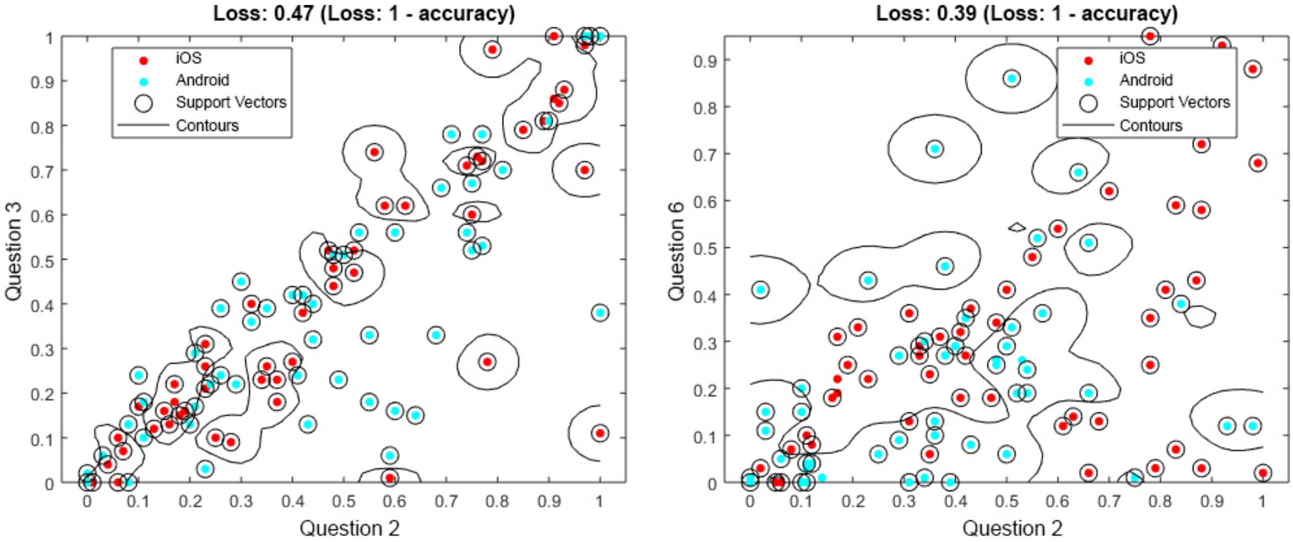
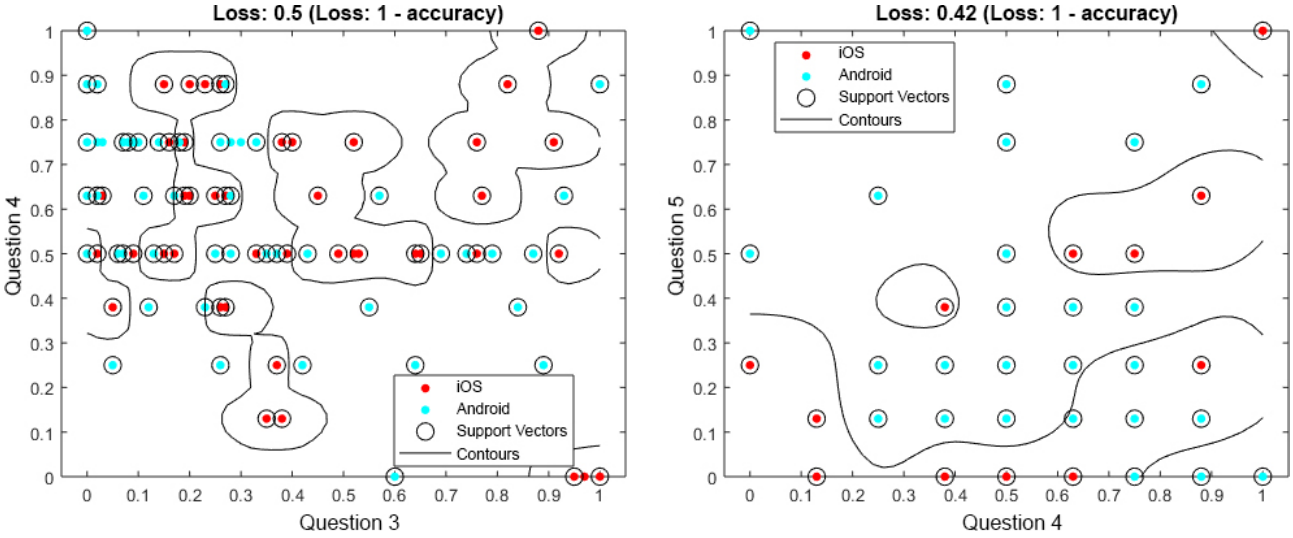


Figure 7. Support vector machine (SVM) results for question 4 (Q4), combined with question 3 (Q3) and question 5 (Q5), from 100 data entries. Q3: How stressful is the tinnitus right now? (slider); Q4: How is your mood right now? (manikins); and Q5: How is your arousal right now? (manikins).



Discussion

Principal Findings

This study evaluated whether it is possible to predict the mobile OS used by respondents for a provided EMA-D questionnaire based on the eight daily tinnitus questions included in the questionnaire, which was administered via TYT. Although the applied machine learning approaches showed different prediction results, in general, the achieved accuracies indicate that the mobile OS is a confounder that must be further considered. This confirms the investigated research question. We are able to predict the mobile OS used with high accuracy based on the dynamic daily assessment data. Compared to Pryss et al [12], the users' ages were no longer different between Android and iOS users, which might be explained by the selection of the sample for this study: only users with more than 10 completed EMA-D questionnaires were selected. In addition to our prior works [12,14], this study shows that the mobile OS not only reveals insights into the tinnitus characteristics of the users, but it is possible to predict the mobile OS based on the provided daily TYT data. On top of this, widely used machine learning approaches with commonly used frameworks and without parameter tuning are able to predict the mobile OS with high accuracy. Note that the RFC achieved the highest prediction result of 78.94%, with default parameter settings using Python scikit-learn. In this context, question 7—How much did you concentrate on the things you are doing right now? (slider)—of the EMA-D questionnaire, which measures the concentration level of a TYT user at the moment, has especially revealed a high accuracy for the RFC prediction. In summary, four important results were found. First, the research question can be answered positively. We are able to predict the mobile OS used for a given EMA-D questionnaire with high accuracy using machine learning methods. Second, the prediction is possible with well-known machine learning methods and frameworks without parameter tuning. Third, machine learning indicates promising results on the EMA-D from TYT users. Therefore, this result should be exploited for further analyses. Fourth, when using mobile devices to collect clinically relevant data, the mobile OS used might be a confounder. Therefore, this information should be collected for each measurement and could be a relevant covariate in data analyses.

Strengths and Limitations

In general, as a strength of this work, it could be shown that the technical peculiarities of different mobile OS types must be considered for the collection of clinically relevant data. As another positive aspect of this work, it could be shown that the types of answers for the questions do not necessarily indicate that a particular answer type, such as a slider, is used a priori with a bias. Otherwise, sliders or any other answer type would be more important than others. In general, we aimed at technically implementing TYT in a way that made sure the questionnaires looked identical on Android and iOS devices as well as having no default setting [11,28,29]. Despite this way of implementing the questionnaires visually, a potential bias cannot be excluded. Therefore, further investigations are required. For example, the sliders on Android and iOS have different numbers of decimal places. On Android, only 2 decimal places are stored, while on iOS, more than 2 decimal places are stored. For the investigation in this paper, the scales of all sliders were harmonized, but such differences must also be further investigated.

For the aspect of whether EMA-D can be used to predict not only the TYT assessments but the TYT users in general, we are conducting another study, in which we investigate whether we are able to predict the mobile OS used on the user level instead of on the assessment level. However, such investigation requires many more considerations. For example, how can we ensure that the training dataset users have similar characteristics as the users for which we apply the trained classifier? Note that such an investigation requires efforts regarding the frameworks used and their provided features.

Conclusions

This work has shown opportunities on one hand and limitations on the other. A particular strength of this study is that TYT has a unique dataset, which is able to comprehensively compare Android and iOS OS types in a medical context. However, the different results between different machine learning approaches showed that it is difficult to predict which questions and answer types are, in general, appropriate for predictions. If a new platform shall be realized and one goal of the platform constitutes using machine learning methods for a prediction, this analysis has not revealed general guidelines that can be followed. Thus, these results can only be seen as a particular outcome for TYT. In addition, when gathering additional contextual information from the TYT users, such as geospatial data, new investigations become possible. In a recent work [27], for example, we investigated geospatial data of mobile crowdsensing users and whether their movement behavior could be a predictor for their current stress situation. As this work also revealed promising results, in the next version of TYT, GPS data can be gathered while filling out the EMA-D questionnaire, if a user allows this measurement.

In future work, we will further address the following three aspects. First, more studies must confirm the results of this work. Second, the results of TYT must be compared to other similar EMA datasets in order to confirm the results between different scenarios. Third, we need to conduct this study again based on the user level instead of on the assessment level.

However, if future work can confirm the presented results, then the combination of EMA, mobile crowdsensing, and machine learning seems to be a worthwhile research endeavor. Nevertheless, we are far from using the results of this work in clinical practice. On the other hand, together with already-revealed medical insights on TYT [6,37-40], the results of this work show that new opportunities are possible in the broader EMA and mobile crowdsensing contexts. In particular, EMA data that were gathered by mobile devices, as well as the crowdsensing paradigm, seem to be promising targets for the application of machine learning algorithms.

Acknowledgments

This work was supported by the University of Würzburg within the funding program Open Access Publishing.

Authors' Contributions

RP substantially contributed to the TYT platform, study design, data analysis, and data interpretation and wrote as well as revised the manuscript. WS substantially contributed to the TYT platform and data interpretation and revised the manuscript. BH substantially contributed to the TYT platform, study design, data analysis, and data interpretation and drafted as well as revised the manuscript. MR substantially contributed to the TYT platform and revised the manuscript. MS substantially contributed to data interpretation and revised the manuscript. BL substantially contributed to data interpretation and revised the manuscript. MB substantially contributed to the data analysis and data interpretation and revised the manuscript. TP substantially contributed to the TYT platform, study design, data analysis, and data interpretation and wrote as well as revised the manuscript.

Conflicts of Interest

None declared.

References

1. Kraft R, Schlee W, Stach M, Reichert M, Langguth B, Baumeister H, et al. Combining mobile crowdsensing and ecological momentary assessments in the healthcare domain. Front Neurosci 2020;14:164 [FREE Full text] [doi: 10.3389/fnins.2020.00164] [Medline: 32184708]
2. Baksa D, Gecse K, Kumar S, Toth Z, Gal Z, Gonda X, et al. Circadian variation of migraine attack onset: A review of clinical studies. Biomed Res Int 2019;2019:4616417 [FREE Full text] [doi: 10.1155/2019/4616417] [Medline: 31534960]
3. Menictas M, Rabbi M, Klasnja P, Murphy S. Artificial intelligence decision-making in mobile health. Biochem (Lond) 2019;41(5):20. [doi: 10.1042/BIO04105020]
4. Statcounter Global Stats. Mobile operating system market share worldwide: May 2019 - May 2020 URL: https://gs.statcounter.com/os-market-share/mobile/worldwide [accessed 2020-06-03]
5. Langguth B. A review of tinnitus symptoms beyond 'ringing in the ears': A call to action. Curr Med Res Opin 2011 Aug;27(8):1635-1643. [doi: 10.1185/03007995.2011.595781] [Medline: 21699365]
6. Schlee W, Pryss RC, Probst T, Schobel J, Bachmeier A, Reichert M, et al. Measuring the moment-to-moment variability of tinnitus: The TrackYourTinnitus smart phone app. Front Aging Neurosci 2016;8:294 [FREE Full text] [doi: 10.3389/fnagi.2016.00294] [Medline: 28018210]
7. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
8. Guo B, Wang Z, Yu Z, Wang Y, Yen N, Huang R, et al. Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. ACM Comput Surv 2015 Sep 29;48(1):1-31. [doi: 10.1145/2794400]
9. Estellés-Arolas E, González-Ladrón-de-Guevara F. Towards an integrated crowdsourcing definition. J Inf Sci 2012 Mar 09;38(2):189-200. [doi: 10.1177/0165551512437638]
10. Pryss R. Mobile crowdsensing in healthcare scenarios: Taxonomy, conceptual pillars, smart mobile crowdsensing services. In: Baumeister H, Montag C, editors. Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics. Cham, Switzerland: Springer International Publishing; Nov 01, 2019:221-234.
11. Pryss R, Probst T, Schlee W, Schobel J, Langguth B, Neff P, et al. Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the TrackYourTinnitus mobile platform. Int J Data Sci Anal 2018 Mar 12;8(4):327-338. [doi: 10.1007/s41060-018-0111-4]
12. Pryss R, Reichert M, Schlee W, Spiliopoulou M, Langguth B, Probst T. Differences between Android and iOS users of the TrackYourTinnitus mobile crowdsensing mHealth platform. In: Proceedings of the 31st International Symposium on Computer-Based Medical Systems. New York, NY: IEEE; 2018 Jul 23 Presented at: 31st International Symposium on Computer-Based Medical Systems; June 18-21, 2018; Karlstad, Sweden p. 411-416 URL: https://ieeexplore.ieee.org/abstract/document/8417273/authors#authors [doi: 10.1109/CBMS.2018.00078]
13. Ubhi HK, Kotz D, Michie S, van Schayck OCP, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171 [FREE Full text] [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]
14. Pryss R, Schlee W, Reichert M, Kurthen I, Giroud N, Jagoda L, et al. Ecological momentary assessment-based differences between Android and iOS users of the TrackYourHearing mHealth crowdsensing platform. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. New York, NY: IEEE; 2019 Oct 07 Presented at: Proceedings Annual International Conference of the IEEE Engineering in Medicine and Biology Society; July 23-27, 2019; Berlin, Germany p. 3951-3955 URL: https://ieeexplore.ieee.org/abstract/document/8857854 [doi: 10.1109/embc.2019.8857854]
15. Mohr DC, Zhang M, Schueller SM. Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
16. Boyer EW, Smelson D, Fletcher R, Ziedonis D, Picard RW. Wireless technologies, ubiquitous computing and mobile health: Application to drug abuse treatment and compliance with HIV therapies. J Med Toxicol 2010 Jun;6(2):212-216 [FREE Full text] [doi: 10.1007/s13181-010-0080-z] [Medline: 20623215]
17. Colombo D, Palacios AG, Alvarez JF, Patané A, Semonella M, Cipresso P, et al. Current state and future directions of technology-based ecological momentary assessments and interventions for major depressive disorder: Protocol for a systematic review. Syst Rev 2018 Dec 13;7(1):233 [FREE Full text] [doi: 10.1186/s13643-018-0899-y] [Medline: 30545415]
18. Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, et al. Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. Curr Psychiatry Rep 2018 Jun 28;20(7):51. [doi: 10.1007/s11920-018-0914-y] [Medline: 29956120]
19. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 2018 May 07;14:91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]

20. Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H. Using mobile phone sensor technology for mental health research: Integrated analysis to identify hidden challenges and potential solutions. J Med Internet Res 2018 Jul 30;20(7):e10131 [FREE Full text] [doi: 10.2196/10131] [Medline: 30061092]
21. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
22. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
23. Delgadillo J. ISRCTN Registry. 2018. StratCare trial: Pragmatic randomised controlled trial of a stratified care model for depression and anxiety URL: http://www.isrctn.com/ISRCTN11106183 [accessed 2020-06-03]
24. Boukhechba M, Cai L, Chow P, Fua K, Gerber M, Teachman B, et al. Contextual analysis to understand compliance with smartphone-based ecological momentary assessment. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY: Association for Computing Machinery (ACM); 2018 Presented at: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018; New York, NY p. 232-238 URL: https://dl.acm.org/doi/10.1145/3240925.3240967 [doi: 10.1145/3240925.3240967]
25. Muniandi LP, Schlee W, Pryss R, Reichert M, Schobel J, Kraft R, et al. Finding tinnitus patients with similar evolution of their ecological momentary assessments. In: Proceedings of the 31st International Symposium on Computer-Based Medical Systems. New York, NY: IEEE; 2018 Presented at: 31st International Symposium on Computer-Based Medical Systems; June 18-21, 2018; Karlstad, Sweden p. 112-117 URL: https://ieeexplore.ieee.org/document/8417222 [doi: 10.1109/cbms.2018.00027]
26. Unnikrishnan V, Beyer C, Matuszyk P, Niemann U, Pryss R, Schlee W, et al. Entity-level stream classification: Exploiting entity similarity to label the future observations referring to an entity. Int J Data Sci Anal 2019 Feb 22;9(1):1-15. [doi: 10.1007/s41060-019-00177-1]
27. Pryss R, John D, Reichert M, Hoppenstedt B, Schmid L, Schlee W, et al. Machine learning findings on geospatial data of users from the TrackYourStress mHealth crowdsensing platform. In: Proceedings of the 20th International Conference on Information Reuse and Integration for Data Science. New York, NY: IEEE; 2019 Presented at: 20th International Conference on Information Reuse and Integration for Data Science; July 30-August 1, 2019; Los Angeles, CA p. 350-355 URL: https://ieeexplore.ieee.org/abstract/document/8843454 [doi: 10.1109/iri.2019.00061]
28. Pryss R, Reichert M, Langguth B, Schlee W. Mobile crowd sensing services for tinnitus assessment, therapy, and research. In: Proceedings of the 3rd International Conference on Mobile Services. New York, NY: IEEE; 2015 Presented at: 3rd International Conference on Mobile Services; June 27-July 2, 2015; New York, NY p. 352-359 URL: https://ieeexplore.ieee.org/abstract/document/7226711 [doi: 10.1109/mobserv.2015.55]
29. Pryss R, Schlee W, Langguth B, Reichert M. Mobile crowdsensing services for tinnitus assessment and patient feedback. In: Proceedings of the International Conference on AI & Mobile Services (AIMS). New York, NY: IEEE; 2017 Presented at: International Conference on AI & Mobile Services (AIMS); June 25-30, 2017; Honolulu, HI p. 22-29 URL: https://ieeexplore.ieee.org/abstract/document/8027265 [doi: 10.1109/aims.2017.12]
30. Xiong H, Huang Y, Barnes L, Gerber M. Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, NY: Association for Computing Machinery (ACM); 2016 Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12-16, 2016; Heidelberg, Germany p. 415-426 URL: https://dl.acm.org/doi/10.1145/2971648.2971711 [doi: 10.1145/2971648.2971711]
31. Chow P, Bonelli W, Huang Y, Fua K, Teachman BA, Barnes LE. DEMONS: An integrated framework for examining associations between physiology and self-reported affect tied to depressive symptoms. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. New York, NY: Association for Computing Machinery (ACM); 2016 Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12-16, 2016; Heidelberg, Germany p. 1139-1143 URL: https://dl.acm.org/doi/10.1145/2968219.2968300 [doi: 10.1145/2968219.2968300]
32. Beierle F, Tran VT, Allemand M, Neff P, Schlee W, Probst T, et al. What data are smartphone users willing to share with researchers? J Ambient Intell Humaniz Comput 2019 Jun 20;11(6):2277-2289. [doi: 10.1007/s12652-019-01355-6]
33. scikit-learn. sklearn.utils.random.sample_without_replacement URL: https://scikit-learn.org/stable/modules/generated/sklearn.utils.random.sample_without_replacement.html [accessed 2020-06-08]
34. Santhanam T, Padmavathi M. Application of k-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. Procedia Comput Sci 2015;47:76-83. [doi: 10.1016/j.procs.2015.03.185]
35. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]

36. Percha B, Baskerville EB, Johnson M, Dudley JT, Zimmerman N. Designing robust N-of-1 studies for precision medicine: Simulation study and design recommendations. J Med Internet Res 2019 Apr 01;21(4):e12641 [FREE Full text] [doi: 10.2196/12641] [Medline: 30932871]
37. Probst T, Pryss R, Langguth B, Schlee W. Emotional states as mediators between tinnitus loudness and tinnitus distress in daily life: Results from the "TrackYourTinnitus" application. Sci Rep 2016 Feb 08;6:20382 [FREE Full text] [doi: 10.1038/srep20382] [Medline: 26853815]
38. Schlee W, Kraft R, Schobel J, Langguth B, Probst T, Neff P, et al. Momentary assessment of tinnitus—How smart mobile applications advance our understanding of tinnitus. In: Baumeister H, Montag C, editors. Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics. Cham, Switzerland: Springer International Publishing; Nov 01, 2019:209-220.
39. Probst T, Pryss R, Langguth B, Schlee W. Emotion dynamics and tinnitus: Daily life data from the "TrackYourTinnitus" application. Sci Rep 2016 Aug 04;6:31166 [FREE Full text] [doi: 10.1038/srep31166] [Medline: 27488227]
40. Probst T, Pryss RC, Langguth B, Rauschecker JP, Schobel J, Reichert M, et al. Does tinnitus depend on time-of-day? An ecological momentary assessment study with the "TrackYourTinnitus" application. Front Aging Neurosci 2017;9:253 [FREE Full text] [doi: 10.3389/fnagi.2017.00253] [Medline: 28824415]

Abbreviations

CSV: comma-separated values
DT: decision tree
EMA: ecological momentary assessment
EMA-D: dynamic ecological momentary assessment
EMA-S: static ecological momentary assessment
FNN: feedforward neural network
mHealth: mobile health
OS: operating system
RFC: random forest classifier
SF28: SmokeFree28
SVM: support vector machine
TYH: TrackYourHearing
TYT: TrackYourTinnitus

Edited by G Eysenbach; submitted 18.07.19; peer-reviewed by A González, EM Messner, M Zolnoori; comments to author 21.10.19; revised version received 23.12.19; accepted 29.02.20; published 30.06.20.

Please cite as:
Pryss R, Schlee W, Hoppenstedt B, Reichert M, Spiliopoulou M, Langguth B, Breitmayer M, Probst T
Applying Machine Learning to Daily-Life Data From the TrackYourTinnitus Mobile Health Crowdsensing Platform to Predict the Mobile Operating System Used With High Accuracy: Longitudinal Observational Study
J Med Internet Res 2020;22(6):e15547
URL: http://www.jmir.org/2020/6/e15547/
doi:10.2196/15547
PMID:32602842

Original Paper

Modifications to Electronic Nicotine Delivery Systems: Content Analysis of YouTube Videos

Zachary B Massey[1], PhD; Yachao Li[2], PhD; Jessica Holli[3], MS; Victoria Churchill[1], MPH; Bo Yang[4], PhD; Katherine Henderson[1], MPH; David L Ashley[1], PhD; Jidong Huang[1], PhD; Lucy Popova[1], PhD

[1]School of Public Health, Georgia State University, Atlanta, GA, United States
[2]Department of Communication Studies and Department of Public Health, The College of New Jersey, Ewing, NJ, United States
[3]Research and Training Institute, John Snow Incorporated, Boston, MA, United States
[4]Department of Communication, University of Arizona, Tuscon, AZ, United States

Corresponding Author:
Lucy Popova, PhD
School of Public Health
Georgia State University
P.O. Box 3995
Atlanta, GA, 30302-3995
United States
Phone: 1 4044139338
Fax: 1 4044131140
Email: lpopova1@gsu.edu

Abstract

Background: As user modification can alter the addictiveness and toxicity of electronic nicotine delivery systems (ENDS), more research is needed to understand the types, motivations, risks, and information sources that lead to these product alterations. YouTube has been identified as a major platform where ENDS users obtain and share information about ENDS products and modifications. However, a comprehensive study of ENDS modification videos on YouTube is lacking.

Objective: This study aimed to analyze the content of YouTube videos depicting modifications of ENDS.

Methods: YouTube was searched in March 2019 to identify videos depicting ENDS modifications. Search terms were derived from interviews with ENDS users and current literature. We used 28 search phrases that combined the words vape and vaping with modification-related key terms (eg, custom build, modification, and dripping). The final sample included 168 videos.

Results: Videos were 1 to 108 min long (median 9.55). Presenters were largely male (117/168, 69.6%), white (94/168, 56.0%), and older than 25 years (94/168, 56.0%). Most videos gave *how to* instructions (148/168, 88.1%), but few offered warnings (30/168, 17.9%) or mentioned commercial alternatives to modifications they presented (16/168, 9.5%). The ENDS devices most often featured were drippers (63/168, 37.5%) and refillable tanks (37/168, 22.0%). The most often modified ENDS components were coils (82/168, 48.8%) and e-liquids (34/168, 20.2%), which included adding other substances, such as cannabis, to the e-liquids (6/168, 3.6%). Most videos portrayed ENDS modifications positively (106/168, 63.1% positive; 60/168, 35.7% neutral; and 2/168, 1.2% negative) and were either neutral or positive in their overall portrayal of ENDS devices (78/168, 46.4% positive; 89/168, 53.0% neutral; and 1/168, 0.6% negative).

Conclusions: This study identified several concerning trends in popular YouTube videos on ENDS modifications, including lack of warnings, the addition of marijuana derivatives to e-liquids, and the positive portrayal of ENDS devices and modifications. By identifying the types of modifications (coil and e-liquid being the most prevalent), this study sets an agenda for research on the effects of modifications.

(J Med Internet Res 2020;22(6):e17104) doi:10.2196/17104

KEYWORDS

ENDS modifications; YouTube; coils; e-liquid; vaping

Introduction

Background

Since the mid-2000s, electronic nicotine delivery systems (ENDS) have become an increasingly popular method for inhaling nicotine from tobacco in the United States [1]. As a category, ENDS include a wide variety of products consisting of a battery, a heater (coil), a mouthpiece, and a chamber containing e-liquid (a solution of propylene glycol [PG] or vegetable glycerine [VG] and other chemicals, usually nicotine and flavors). The heater vaporizes the e-liquid, producing an aerosol that is inhaled. ENDS can also be called vapes, e-hookahs, and e-cigarettes. Since June 2019, an outbreak of lung disease linked to vaping and the use of ENDS has sickened 2506 and killed 54 people [2]. Although this outbreak is ongoing, the US Centers for Disease Control and Prevention urges ENDS users not to modify the products or add substances not intended by the manufacturer, highlighting the potential role ENDS modifications might play in this outbreak [3]. For this study, we define *modification* as both product misuse and tampering unintended by the manufacturers as well as alteration, customization, adjustment, and user choice of e-liquid or accessories made within manufacturer specifications. Examples of ENDS modifications include coil replacement, mixing of e-liquids, or increasing battery voltage. Although some users view the modifiability of ENDS devices as a positive attribute [4,5], research has found that modifications can have harmful effects, such as creating higher levels of toxic emissions when users increase power to the coil [6-8]. Given the popularity of ENDS and health risks from modification, research is needed to understand the primary ways ENDS are modified by users.

To date, little research has described the methods of ENDS modifications. YouTube is the most popular video-sharing site in the world [9], and there are a large number of ENDS videos on YouTube [10], including videos about ENDS modifications [11]. YouTube has been identified as a major platform where ENDS users obtain and share information about ENDS products and modifications [12-14]. However, a comprehensive study of ENDS modification videos on YouTube is lacking. One recent study examined YouTube videos depicting orthodox (intended by manufacturer) and unorthodox (unintended by the manufacturer) ENDS modifications [11]. Results showed that videos depicting unorthodox use were three times more prevalent than videos depicting orthodox use, although the analysis only focused on open devices and e-liquids and not on closed devices [11]. Furthermore, the sample was collected in 2016 and may not account for recent trends in ENDS modification. A more in-depth study is needed to update and assess the breadth of ENDS modifications on YouTube.

Objectives

In this paper, we examined YouTube videos posted between 2013 and 2019 showing modifications of ENDS devices pertaining to hardware (eg, coil, battery, pods, and other modifications) and e-liquids. Our objective was to understand the video features (including presenter characteristics, if applicable) as well as the types, motivations, communication of risks, and information sources of ENDS modifications. Although other studies have examined the social acceptability of ENDS on YouTube [15], unorthodox use of ENDS on YouTube [11], and promotion of vape tricks on YouTube [16], this study, to the best of our knowledge, is the first to analyze modifications to the full spectrum of ENDS devices on YouTube.

Methods

Identifying Electronic Nicotine Delivery Systems Modification Videos

YouTube was searched on March 15, 2019, to identify videos depicting ENDS modifications. An account was created on an incognito (private) browser to ensure browsing history did not influence the results. Search terms were derived from interviews with ENDS users [14] and the current literature [10] indicating that “vape” and “vaping” are common and inclusive terms used by ENDS enthusiasts to describe a number of different ENDS devices. Search phrases included the following: “vape DIY (ie, do it yourself),” “vape build DIY,” “vape custom build,” “vape modification,” “vape rebuild DIY,” “vape coil rebuild,” “vape coil DIY,” “vape ohm rebuild,” “vape ohm DIY,” “vape e-juice DIY,” “vape e-juice custom build,” “vape voltage DIY,” “vape voltage custom build,” “vape RBA (ie, rebuildable atomizer) custom build,” “vape RBA DIY,” “vape dripping custom build,” “vape dripping DIY,” “vape refill custom build,” “vape refill DIY,” “vaping custom build,” “vaping DIY,” “vaping modif* custom build,” “vaping rebuild DIY,” “mods rebuild DIY,” “vape chang* DIY,” “vape chang* custom build,” “vape wattage DIY,” and “vape wattage custom build.” Search terms were entered into the YouTube search engine, and the top 10 results for each phrase were included (N=280). After removing duplicates, non-English videos, and those not presenting ENDS modifications, 207 videos remained. Each video was reviewed by trained coders to ensure that the content included ENDS modifications. This verification process excluded 39 videos, resulting in a final sample of 168 videos. This study was approved by the Georgia State University Institutional Review Board (H19055).

Coding Scheme

We developed a codebook based on interviews with ENDS users [14], previous literature [10,15], and expert consensus. By expert consensus, we refer to review by our interdisciplinary team of collaborators with substantial expertise in ENDS use behavior and ENDS product research. We collected information on the attributes of the videos (eg, length, number of views, number of likes and dislikes, and date of posting). We coded the video source as individual, retailer, manufacturer, or group or organization. We noted whether the videos included links to vendors, whether the videos were followed by other modification videos via YouTube’s auto-play feature (ie, showing content based on user search history), gave *how to* instructions, included a warning (a warning statement about the health risks of smoking or a formal, legal, or conflict of interest disclaimer), or offered information about commercially available alternatives to modifications. If a person appeared in the video, we recorded if it was a single presenter or multiple presenters and their sex (male or female), race (white, black, other, or cannot be determined), and estimated age range (youth: younger than 18 years, young adult: approximately 18-25 years, or adult: older than 25 years). We classified the device(s) featured in the video as dripper, refillable tank, closed pod, refillable pod, home-built refillable tank, cig-a-like, squonk, 2-pod systems, and stem/salt hybrid [17]. Definitions for ENDS devices are provided in Table 1. If multiple devices were shown in a single video, all devices, regardless of their prominence, were included in the coding.

Table 1. Content codes for electronic nicotine delivery systems devices (N=168) and the percentage and number of videos in the final sample that featured each device type.

Device type	Definition	Value, n (%)
Dripper	These devices do not include a tank but do include a tip, coil, wicks, battery housing, battery, and fire button. There is no e-liquid reservoir—e-liquid is squirted directly on to the coil housing or *deck* and heated up as soon as the fire button is pressed. They are intended to be taken apart and put back together, which makes them an open system. Some or most of the device has been hand-built, with some parts being retrofitted from existing commercially produced devices.	63 (37.5)
Refillable tank	A commercially produced refillable tank system, which features, at minimum, a tip, a glass tank, a coil, a battery unit, and a fire button. The tip, tank, coil, and batteries can be removed, and the tip, tank, coil, and battery unit can be easily separated from each other and reconnected, which makes it an open device. For commercially produced devices, the battery unit is shaped either like a rectangular box (box mod) or like a cylinder (vape pen).	37 (22.0)
Closed pod	A commercially produced closed pod system with a single pod that cannot be refilled, such as a Juul. These are typically small, thin, rectangular boxes about the shape of a 4-inch long USB drive. However, some may also be parallelogram, square, or teardrop shaped.	7 (4.2)
Refillable pod	A commercially produced closed pod system with a single pod that can be refilled. These are similar to closed pod systems, except that the pods include a small port that allows the e-liquid reservoir to be refilled. Refillable pod devices with only 1 pod usually use salt nicotine (higher concentration and lower wattage).	5 (3.0)
Home-built refillable tank	A home-built device featuring a refillable tank that is not commercially produced. They are intended to be taken apart and put back together, which makes them an open system. Some or most of the device has been hand-built, with some parts being retrofitted from existing commercially produced devices. Home-built parts may include a tip, battery housing, and fire button. Retrofitted parts may include a tip, tank, coil, electronics, and batteries.	4 (2.4)
Cig-a-like	A commercially produced device designed to look like a combustible cigarette in size, shape, and color (sometimes including a brown/tan filter tip). However, it is entirely electronic and includes a battery, a small e-liquid reservoir, and an atomizer. Some are not rechargeable, nor can they be taken apart but are intended to be disposed of once the e-liquid has been used up. Others are rechargeable and have replaceable pods.	4 (2.4)
Squonk	A squonker, also known as squonk mod, is a type of dripper. Like drippers, they include a tip, coil, wicks, battery housing, battery, and fire button. There is, however, a small e-liquid reservoir made out of a pliable material such as silicone, which can be refilled. The e-liquid reservoir is accessible through a hole or window in the device, which allows the reservoir to be compressed. Compressing the reservoir squirts e-liquid directly on the coil, which heats it up as soon as the fire button is pressed. Squonks are only commercially produced and typically look like small box mods without the tank on top of the mod, but with the e-liquid reservoir visible and accessible through a small window on the side.	4 (2.4)
2 pod systems	Devices that include 2 pods of the same type and 2 receptors that can be swapped. This allows the device to provide vapor with 2 distinct flavors as well as vapor that blends the 2 flavors.	0 (0.0)
Stem/salt hybrid	A hybrid that includes 2 refillable pods each with its own receptor on the device. One is intended for salt nicotine e-liquid (higher concentration/lower wattage) and the other is for *stem* nicotine, that is, typical e-liquid (lower concentration of nicotine, if at all/higher wattage). Each pod can only fit into 1 receptor on the device, and they cannot be swapped with the other receptor.	0 (0.0)
Multiple devices	Several devices are featured.	7 (4.2)
Other	Device featured is another product not listed or unsure of the product type.	15 (8.9)

We developed a coding scheme to capture ENDS modifications to the coil, e-liquid, battery, pods, and other modifications. For each type of modification, we developed codes for common reasons for this modification, including an *other* option with text descriptions. Each modification was coded as 1 (present) or 0 (absent). The detailed codes with descriptions for modifications and reasons are presented in Table 2.

Finally, we coded the video's tone toward the featured ENDS modification and ENDS devices in general as positive, neutral, or negative.

Table 2. Codes and definitions of modifications and reasons for modifications (N=168).

Code	Definition	Value[a], n (%)
Coil modification		
Coil build	Building the coil	60 (35.7)
Replacing coil	Replacing existing coil with a different coil from the manufacturer	9 (5.4)
Coil voltage	Altering the voltage of the existing coil through controls built into the device by the manufacturer	8 (4.8)
Number of coils	Altering the number of coils in the device	1 (0.6)
Coil gauge	Explicitly replacing an existing coil with a coil that is a different gauge wire	3 (1.8)
Pod coil	Cleaning the coils built into the pods/closed systems	2 (1.2)
Other coil modification	Other type of coil modification	35 (20.8)
Reasons for coil modification		
Increased voltage	To increase the voltage of the heating coil	9 (5.4)
Other	Other reasons for coil modification	31 (18.5)
E-liquid modification		
E-liquid mixing	Mixing e-liquid includes making your own e-liquid from standard components (any of PG[b], VG[c], nicotine, flavors, etc) or altering bought e-liquid with any of these components (adding flavors and changing PG/VG ratio)	20 (11.9)
E-liquid concentration	Replacing e-liquids with different levels of nicotine concentrations	3 (1.8)
Adding substances	Adding other substance (eg, cannabis oil) or replacing e-liquids with other substances	6 (3.6)
Other e-liquid modification	Other type of e-liquid modification	13 (7.7)
Reasons for e-liquid modification		
Flavor	To enhance flavors in e-liquid	16 (9.5)
Cost	To save money/reduce the cost of manufactured liquid	10 (6.0)
Other	Other reasons for e-liquid modification	15 (8.9)
Battery modification		
Replace battery	Replacing a battery with a battery with different properties	0 (0.0)
Battery configuration	Changing battery configuration, either in series or in parallel	0 (0.0)
Additional battery	Attaching additional batteries	0 (0.0)
Other battery modification	Other modification to battery	13 (7.7)
Reasons for battery modification		
Increased power	To increase the power of the existing tank device	0 (0.0)
Other	Other reasons for battery modification	4 (2.4)
Pod modification		
Pod refilling	Refilling pods with e-liquids	9 (5.4)
Reasons for pod modification		
Extend life	To extend the life of a pod/cartridge	4 (2.4)
Other	Other reasons of modifications to closed pod	7 (4.2)
Other modification		
Modifying controls	Modifying software regulating controls	2 (1.2)
Improved features	Replacing existing device with another device with improved features	1 (0.6)
Additional features	Replacing existing device with another device with additional features	3 (1.8)
Other	Other types of modification that are not listed above	22 (13.1)

[a] The number of videos and the percentage in the final sample with each code present.

[b] PG: propylene glycol.

[c] VG: vegetable glycerin.

Video Coding

The coding team consisted of 5 members, who received a minimum of four 2-hour training sessions. The coders coded 5 to 10 videos for each training session, and discrepancies were discussed and resolved. Training videos were not included in the final data analysis because the code book was further revised during the training. To establish intercoder reliability, approximately 15% of videos (25/168, 14.9%) were randomly selected from the final sample. Due to the large amount of time spent to code each video, sequential-overlapping coding was used for reliability testing [18]. Specifically, each of the 25 videos was coded by 2 coders, and each coder coded 10 videos in total. Using Randolph's free-marginal kappa [19,20], the intercoder reliability of each variable was high, ranging from 0.74 to 1.0. The remaining videos were then divided and independently coded by 4 coders.

Data Analysis

Data were analyzed using IBM SPSS 25. Descriptive statistics were performed to assess the frequency of each coding variable.

Results

Characteristics of Videos, Presenters, and Device Types

Videos (N=168) ranged from 1 to 108 min in length (median 9.55) and accounted for a total of 112,043,718 views (median 168,234) as of March 15, 2019. Videos had more *likes* than *dislikes* (12 to 1 ratio). Examining videos by year showed an increase from 2013 (2/168, 1.2%) to 2014 (19/168, 11.3%), to 2015 (27/168, 16.1%), to 2016 (34/168, 20.2%), peaking in 2017 (51/168, 30.4%), and decreasing from 2018 (29/168, 17.3%) to 2019 (6/168, 3.6%). The most common source for videos were individuals (135/168, 80.4%), followed by retailers (18/168, 10.7%), manufacturers (8/168, 4.8%), and groups/organizations (7/168, 4.2%). A majority of videos (120/168, 71.4%) provided a link to a vendor's site, and most videos (136/168, 81.0%) were followed by another modification video via auto-play. Most videos gave *how to* instructions (148/168, 88.1%); few videos offered warnings (30/168, 17.9%) or mentioned commercially available alternatives (16/168, 9.5%).

Most videos showed 1 or more persons (125/168, 74.5%), and presenters were largely male (117/168, 69.6%), white (94/168, 56.0%), and appeared to be adults older than 25 years (94/168, 56.0%). Modifications were most frequently performed on drippers (63/168, 37.5%) and refillable tanks (37/168, 22.0%; Table 3).

Table 3. Modification by device type.

Device type	Coil[a], n	E-liquid[a], n	Battery[a], n	Pods[a], n	Other[a], n
Dripper	38	7	6	0	3
Refillable tank	28	3	2	0	3
Home-built refillable tank	0	0	2	0	1
Closed pod	0	4	0	6	0
Refillable pod	0	2	0	1	0
Cig-a-like	0	4	0	2	0
Squonk	2	0	0	0	3
Stem/salt hybrid	0	0	0	0	0
Two-pod system	0	0	0	0	0
Multiple devices	3	1	1	0	3
Other	5	1	2	0	9
Total	76	22	13	9	22

[a] Each cell shows the number of videos with each modification. If a video showed a specific modification, the video was coded as 1. Videos coded as 1 include single modifications (eg, building a coil from scratch) or multiple modifications (eg, removing coil, cleaning coil, and rewrapping coil).

Coil Modifications

Modifications to the coil were the most frequently portrayed; building coils from scratch was featured in 35.7% (60/168) of the videos. Coil replacement with a manufactured alternative was depicted in 5.4% (9/168) of videos, and 4.8% (8/168) of the videos showed how to alter coil voltage through controls in the device. Less than 0.6% (1/168) of the videos gave instructions on how to alter the number of coils in the device, and 1.8% (3/168) showed how to replace coils with a wire of different gauge. Only 1.2% (2/168) of the videos showed how to clean the coils in a nonmodifiable pod.

Reasons for modifying in many of the coil videos (31/168, 18.5%) were not in the prespecified list. Among these reasons, building coils from scratch (17/168, 10.1%) and enhancing flavor (6/168, 3.6%) were the most commonly described.

Building coils from scratch included home builds (necessitating a new coil) and adding coils in store-bought devices. According to the presenters, enhancing flavor can be achieved by changing the wire gauge (eg, 28 vs 24 gauge), wire composition (eg, nickel vs stainless steel), or number of wire wraps. Increasing voltage was also listed as a reason for modifying coils (9/168, 5.4%).

The most viewed coil modification video (3,017,588 views) shows a presenter instructing the audience on how to build a sleeper coil (ie, double-wrapped coil made from a single wire). The tutorial includes instructions on how to wrap the coil and attach it to the battery posts. The *how to* portion is book-ended by discussion and demonstration of the modification. Before the tutorial, the presenter describes how the sleeper coil heats e-liquid efficiently, producing a greater cloud and better flavor. After the tutorial, the presenter vapes to demonstrate the size of his aerosol cloud, presumably the result of the sleeper coil.

Modifications of E-Liquids

Modifications of e-liquids were depicted in 20.2% (34/168) of videos. Many of the e-liquid modification videos showed mixing (20/168, 11.9%; eg, mixing store-bought glycol). Separately, 1.8% (3/168) of videos included mixing liquids to increase nicotine levels. Adding other substances (eg, cannabis oil) was found in 3.6% (6/168) of videos.

Enhancing flavor (16/168, 9.5%) and saving money (10/168, 6.0%) were reasons given for e-liquid modifications, as were *other* reasons (15/168, 8.9%). Descriptions of other reasons included creating better clouds, making nicotine-free liquid, and creating marijuana e-liquid. Overall, e-liquid modifications centered on allowing users to control the flavor, price, and composition of e-liquids.

The most viewed e-liquid mixing video (1,769,171 views) shows how to make e-liquids with store-bought materials. The presenter's face is not shown, although the video is narrated. The presenter describes 4 components necessary to make e-liquid: VG, PG, premixed liquid nicotine (75% VG and 25% PG), and store-bought flavoring (eg, strawberry and banana). Before mixing the components, the presenter uses a web-based e-liquid calculator. The calculator uses batch size, nicotine strength (mg/ml), and VG/PG ratio to calculate the mixing recipe. However, the calculator does not account for the addition of flavor. Therefore, the presenter explains the conversion necessary to add these liquids. At the completion of the conversion, the presenter combines the liquids in a cylinder and then attaches a clamp to a power drill, wraps the clamp fingers around the cylinder, and engages the drill, spinning the cylinder and mixing the liquids within. The liquid is then *steeped* in warm water to allow the ingredients to blend together. The video closes by encouraging viewers to document mixes in a spreadsheet.

The most viewed *adding substances video* (717,781 views) gives a tutorial on how to make *weed e-cig juice*. The presenter bakes marijuana on a cooking sheet, breaks it apart by hand, and mixes it with VG. After 3 months of iteratively adding glycerin and remixing the solution, the presenter strains marijuana-infused glycerin through a mesh and then injects the liquid into an ENDS tank. The video closes with the presenter blowing vapor clouds over an advertisement for marijuana accessories. Among the *adding substances* videos, most (5 of 6) focused on adding marijuana derivatives to e-liquids.

Battery Modifications

No videos showed modifications to increase battery power, change configurations, or add additional batteries. The only videos coded for battery modifications were those in the *other battery modification* category (13/168, 7.7%), which showed variations of adding batteries to home-built devices. Examining descriptions for the *other* category showed that all entries were home builds, either adapting household items (eg, a flashlight) or using commercially made parts. Adapting household items to create a vape involved a premade battery, whereas building from scratch sometimes required wiring batteries. This distinction could indicate differences in battery modification hazards.

The most viewed (1,645,630 views) battery modification video transforms a flashlight into an ENDS device. The presenter does not address the audience, but text is displayed to highlight important steps. In the video, the presenter disassembles a flashlight and uses the parts as well as household materials to create a small vape. The body of the flashlight is retrofitted to hold a battery and the head with an atomizer (ie, a coil and wick attached to battery posts housed on a deck). The flashlight button engages the battery, heating the coil and producing a stream of vapor from the atomizer. The battery does not require wiring and appears to be a standard battery consistent with those used in any flashlight.

Other battery videos, however, depict more advanced constructions. One example (763,291 views) shows how to build a *box mod*, which is a highly modifiable ENDS device characterized by a removable atomizer, open tank, and accessible battery housing. The presenter walks the audience through the build, which includes wiring of 2 lithium polymer batteries (ie, LiPo) into the device. LiPos are lightweight batteries that can explode if they are overcharged. Such explosions have been linked to injury of ENDS users [21]. The presenter warns viewers not to let battery wires touch during construction. However, the breadth of the hazard is not disclosed, and the video contains no warnings about the danger of explosions.

Modifications of Pods and Cartridges

A pod is a cartridge containing e-liquids to be inserted in an ENDS device. When the liquid is gone, many manufacturers intend that the cartridge is discarded and not reused (although there are some refillable pods on the market). Refilling nonreusable pods with e-liquid was found in 5.4% (9/168) of videos. The primary reasons for pod modifications (7/168, 4.2%) were not in the specified list, and thus, they were coded as other. Descriptions of *other* reasons included customizing the look of devices, increasing vapor production, and enhancing flavor. Extending the life of a pod was listed as a reason for 2.4% (4/168) of videos.

An example of pod modification (182,621 views) depicts a presenter changing the liquid in a nonreusable pod. In the video, the presenter uses scissors to open the cartridge. The presenter pours the liquid into another container, cleans the pod, and then refills it with an e-liquid called *Heisenberg* (fruit-flavored menthol). After the refill, the presenter takes several drags until the tobacco flavor is gone, and only Heisenberg remains. In this case, the pod modification is motivated by the choice of the flavor.

Other Modifications

Only 1.2% (2/168) of videos showed modifications to the software regulating built-in controls. Replacement of an existing device with another with improved features was found in 0.6% (1/168) of videos, whereas replacing an existing device with another device with additional features was featured in 1.8% (3/168) of videos. Although drippers were the most featured devices (63/168, 37.5%), our analyses did not show any instances of converting nondripping devices to drippers.

Finally, 13.1% (22/168) of videos did not fit within our specified coding categories and were coded as *other* modifications. These videos showed building ENDS as the primary activity. Several videos depicted unusual builds, such as creating a vape from an empty mint box or adapting a soda can into an e-liquid tank. One unique example (66,573 views) showed how to build an ENDS in the dashboard of a car. Other videos showed how to make *drip tips* to replace lost mouthpieces for drippers. In one example (85,006 views), the mouthpiece was made from a copper plumbing fitting. Together, videos coded as *other* showed unique modification behaviors mostly related to building from scratch.

Tone

Most videos used a positive tone to portray ENDS modifications (106/168, 63.1% positive; 60/168, 35.7% neutral; and 2/168, 1.2% negative) but were either neutral or positive in their portrayal of ENDS devices (78/168, 46.4% positive; 89/168, 53.0% neutral; and 1/168, 0.6% negative).

Discussion

Principal Findings

This study provides information about the ways in which users modify ENDS devices on YouTube. Results identified several concerning trends, including lack of warnings, addition of marijuana derivatives to e-liquids, and the positive portrayal of ENDS devices and modifications. Regarding specific modifications, the most often modified components were coils and e-liquids. Modifications to the wick and battery were less often discussed. The sample of modification videos did not depict potentially harmful practices identified in past literature, such as modifying a device for *dripping* or increasing battery power. Together, these results help inform public health researchers by identifying trends in ENDS modifications on YouTube, the most popular video-sharing website in the world [9] and a primary source of information on ENDS devices and modifications [11,15,16].

Our findings identified several concerning trends in modification videos. Specifically, the majority of videos gave *how to* instructions without warning viewers about potential dangers (eg, exploding batteries or lung exposure to harmful constituents). The majority also failed to mention commercially available alternatives, which could be less dangerous than home-grown modifications. ENDS devices and modifications were also depicted in a favorable light, and a majority of videos were followed by another modification video, likely the result of YouTube's default auto-play feature, which queues similar content based on search history. In a situation where YouTube users search for ENDS modifications, favorable portrayals and lack of warnings, coupled with auto-play features, could potentially create a self-reinforcing loop where users search for modification videos. The dangers of modifications are underemphasized to viewers, and these videos are followed by additional videos portraying ENDS in a favorable light, further underemphasizing the dangers of modification. Repeated, user-driven exposure to videos underemphasizing risks could normalize ENDS modifications, which is of particular concern given YouTube's primary audience.

YouTube is the most popular video-sharing site in the world [9], and YouTube has been identified as a major platform where ENDS users obtain and share information about ENDS products and modifications [12-14]. Many YouTube users are young, with 81% of those aged 15 to 25 years in the United States using YouTube [22]. This demographic is often targeted by tobacco companies [23,24] and is a group that often experiments with cigarettes and ENDS [25]. Positive portrayals of modifications without safety warnings, which are algorithmically linked and shown in sequence to impressionable audiences, bear serious consideration from tobacco control experts. Future research may explore how YouTube modification videos could potentially detract from nicotine cessation campaigns aimed at youth and young adults.

Altering the coil was the most prevalent type of modification in our study. Little research has investigated the effects of exposure to the components of coils on user health. One notable study tested e-liquids for neurotoxicants before and after contact with ENDS devices [26]. Results showed higher levels of toxic metals in the aerosol, suggesting that the contamination resulted from the contact of e-liquids with the metals in ENDS devices, including coils. As coil modifications were most prevalent in our analysis, we suggest future research investigating possible health effects of coil modifications.

E-liquid was the second most often modified characteristic in our sample. E-liquid modifications included mixing from scratch and adding other substances. In terms of mixing, most videos cited a desire to control the flavor as a reason for the modification. This finding is informative given the recent interest among policy makers to ban flavored e-liquids from the market [27]. Regulatory policy should consider how regulating ENDS flavors might influence the likelihood that these actions will impact the frequency at which consumers mix their own flavored e-liquids.

Marijuana derivatives were used in most videos depicting substances added to e-liquids. This finding is significant given the recent outbreak of vaping-related lung disease [28]. Early investigation revealed that the majority of patients had a history of vaping tetrahydrocannabinol (THC) products, suggesting that the outbreak may be related to the source of THC or vaping THC in a device intended to be used with nicotine [29]. Given the deadly nature of the new vaping lung disease, this particular modification may demand greater scrutiny from public health experts.

Tampering with devices to expose heating coils for dripping has been cited as a modification trend in past research [30]. Yet, none of the videos in our sample showed how to adapt a nondripping device into a dripper. This finding resonates with recent interviews of ENDS enthusiasts, indicating that dripping modifications may be less frequent due to increased availability of commercially produced drippers and other devices that were previously only available via users' modifications [14]. However, it should be noted that drippers were the most featured devices in YouTube videos. Thus, although dripping conversions were absent in our sample, modifications to drippers were frequent.

Although exploding batteries are a risk [6], we found only a few videos on battery modifications. They were generally associated with building ENDS mods from scratch and showed some dangerous practices (eg, wiring LiPo batteries) that have not been widely discussed previously.

Limitations

As this sample was collected in early 2019, it is unclear if the decline in videos on ENDS modifications we observed since 2017 will continue. Although this decrease corroborates our findings from interviews with ENDS enthusiasts [14], more research is needed to determine if this decrease continues. As this study was conducted before the first cases of e-cigarette– and vaping-associated lung injury were reported, we do not know what impact this outbreak may have had on ENDS modification activities. Our search keywords were not exhaustive and may not capture all ENDS modification videos on YouTube. Due to the variation of ENDS modification videos, our results are descriptive. More research is needed to isolate and identify causal mechanisms that motivate modification behaviors. This work informs future population-level surveillance research aimed at identifying motivations for modifying ENDS devices.

Conclusions

Our content analysis of the full spectrum of ENDS devices and modifications in YouTube videos identified several concerning trends, including lack of warning to viewers, addition of marijuana derivatives to e-liquids, and the positive portrayal of ENDS devices and modifications. At the same time, our analysis did not find certain modifications (eg, dripper conversions or battery tampering), which have been identified previously as a public health concern. By identifying the types of modifications (with modifications to coil and e-liquid being the most prevalent), this study provides a foundation for assessing the prevalence of ENDS modifications in the population and informs the agenda for research assessing the health effects resulting from ENDS modifications.

Acknowledgments

Research reported in this publication was supported by the National Institute of Drug Abuse of the National Institutes of Health and Food and Drug Administration Center for Tobacco Products (R01DA047397). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Food and Drug Administration.

Authors' Contributions

ZBM wrote the first draft. YL and J Holli conducted the data analysis. DLA and LP conceptualized the study and wrote the implications. YL, J Holli, BY, KH, and VC coded videos. All authors contributed to the writing and revision and approved the final version of the manuscript.

Conflicts of Interest

DLA has received funds for work done for the World Health Organization Tobacco Free Initiative, has worked as a consultant for Pfizer and McKing Consulting, Inc, and has worked as a Special Government Employee of the US Food and Drug Administration.

References

1. McMillen RC, Gottlieb MA, Shaefer RM, Winickoff JP, Klein JD. Trends in electronic cigarette use among US adults: use is increasing in both smokers and nonsmokers. Nicotine Tob Res 2015 Oct;17(10):1195-1202. [doi: 10.1093/ntr/ntu213] [Medline: 25381306]
2. Centers for Disease Control and Prevention. 2019. Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html [accessed 2019-10-14]
3. Centers for Disease Control and Prevention. 2020. Smoking & Tobacco Use: Frequently Asked Questions URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease/need-to-know/index.html [accessed 2019-10-14]
4. Bala MM, Strzeszynski L, Topor-Madry R. Mass media interventions for smoking cessation in adults. Cochrane Database Syst Rev 2017 Nov 21;11:CD004704 [FREE Full text] [doi: 10.1002/14651858.CD004704.pub4] [Medline: 29159862]

5. Majmundar A, Kirkpatrick M, Cruz TB, Unger JB, Allem JP. Characterising KandyPens-related posts to Instagram: implications for nicotine and cannabis use. Tob Control 2019 May 30:- epub ahead of print. [doi: 10.1136/tobaccocontrol-2019-055006] [Medline: 31147484]
6. Kosmider L, Sobczak A, Fik M, Knysak J, Zaciera M, Kurek J, et al. Carbonyl compounds in electronic cigarette vapors: effects of nicotine solvent and battery output voltage. Nicotine Tob Res 2014 Oct;16(10):1319-1326 [FREE Full text] [doi: 10.1093/ntr/ntu078] [Medline: 24832759]
7. Sleiman M, Logue JM, Montesinos VN, Russell ML, Litter MI, Gundel LA, et al. Emissions from electronic cigarettes: key parameters affecting the release of harmful chemicals. Environ Sci Technol 2016 Sep 6;50(17):9644-9651. [doi: 10.1021/acs.est.6b01741] [Medline: 27461870]
8. Talih S, Balhas Z, Salman R, Karaoghlanian N, Shihadeh A. 'Direct dripping': a high-temperature, high-formaldehyde emission electronic cigarette use method. Nicotine Tob Res 2016 Apr;18(4):453-459 [FREE Full text] [doi: 10.1093/ntr/ntv080] [Medline: 25863521]
9. Alexa. The Top 500 Sites on the Web URL: https://www.alexa.com/topsites [accessed 2019-10-14]
10. Huang J, Kornfield R, Emery SL. 100 million views of electronic cigarette YouTube videos and counting: quantification, content evaluation, and engagement levels of videos. J Med Internet Res 2016 Mar 18;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
11. Guy MC, Helt J, Palafox S, Green K, Soule EK, Maloney SF, et al. Orthodox and unorthodox uses of electronic cigarettes: a surveillance of YouTube video content. Nicotine Tob Res 2019 Sep 19;21(10):1378-1384 [FREE Full text] [doi: 10.1093/ntr/nty132] [Medline: 29961828]
12. McCausland K, Maycock B, Leaver T, Jancey J. The messages presented in electronic cigarette-related social media promotions and discussion: scoping review. J Med Internet Res 2019 Feb 5;21(2):e11953 [FREE Full text] [doi: 10.2196/11953] [Medline: 30720440]
13. Collins L, Glasser AM, Abudayyeh H, Pearson JL, Villanti AC. E-cigarette marketing and communication: how e-cigarette companies market e-cigarettes and the public engages with e-cigarette information. Nicotine Tob Res 2019 Jan 1;21(1):14-24 [FREE Full text] [doi: 10.1093/ntr/ntx284] [Medline: 29315420]
14. Li Y, Fairman RT, Churchill V, Ashley DL, Popova L. Users' modifications to electronic nicotine delivery systems (ENDS): interviews with ENDS enthusiasts. Int J Environ Res Public Health 2020 Feb 2;17(3):e918 [FREE Full text] [doi: 10.3390/ijerph17030918] [Medline: 32024230]
15. Luo C, Zheng X, Zeng DD, Leischow S. Portrayal of electronic cigarettes on YouTube. BMC Public Health 2014 Oct 3;14:1028 [FREE Full text] [doi: 10.1186/1471-2458-14-1028] [Medline: 25277872]
16. Kong G, LaVallee H, Rams A, Ramamurthi D, Krishnan-Sarin S. Promotion of vape tricks on YouTube: content analysis. J Med Internet Res 2019 Jun 18;21(6):e12709 [FREE Full text] [doi: 10.2196/12709] [Medline: 31215510]
17. Harrell PT, Eissenberg T. Automated dripping devices for vapers: RDTAs, bottomfeeders, squonk mods and dripboxes. Tob Control 2018 Jul;27(4):480-482 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-053817] [Medline: 28735274]
18. Neuendorf KA. The Content Analysis Guidebook. Thousand Oaks, California: Sage Publications; 2016.
19. Randolph JJ. Free-Marginal Multirater Kappa (multirater K[free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. In: Proceedings of the Joensuu Learning and Instruction Symposium. 2005 Presented at: JLIS'05; October 14-15, 2005; Joensuu, Finland URL: https://eric.ed.gov/?id=ED490661
20. Randolph JJ. Justus Randolph. 2008. Online Kappa Calculator URL: http://justusrandolph.net/kappa/ [accessed 2019-10-01]
21. Nicoll KJ, Rose AM, Khan MA, Quaba O, Lowrie AG. Thigh burns from exploding e-cigarette lithium ion batteries: first case series. Burns 2016 Jun;42(4):e42-e46. [doi: 10.1016/j.burns.2016.03.027] [Medline: 27118069]
22. Clement J. Statista. 2019. Percentage of US Internet Users who Use YouTube as of 3rd Quarter 2019, by Age Group URL: https://www.statista.com/statistics/296227/us-youtube-reach-age-gender/ [accessed 2020-02-14]
23. Henriksen L, Schleicher NC, Dauphinee AL, Fortmann SP. Targeted advertising, promotion, and price for menthol cigarettes in California high school neighborhoods. Nicotine Tob Res 2012 Jan;14(1):116-121 [FREE Full text] [doi: 10.1093/ntr/ntr122] [Medline: 21705460]
24. Pierce JP, Messer K, James LE, White MM, Kealey S, Vallone DM, et al. Camel No. 9 cigarette-marketing campaign targeted young teenage girls. Pediatrics 2010 Apr;125(4):619-626. [doi: 10.1542/peds.2009-0607] [Medline: 20231181]
25. Centers for Disease Control and Prevention. 2015. E-Cigarette Use Triples Among Middle and High School Students in Just One Year URL: http://www.cdc.gov/media/releases/2015/p0416-e-cigarette-use.html [accessed 2019-10-14]
26. Olmedo P, Goessler W, Tanda S, Grau-Perez M, Jarmul S, Aherrera A, et al. Metal concentrations in e-cigarette liquid and aerosol samples: the contribution of metallic coils. Environ Health Perspect 2018 Feb 21;126(2):027010 [FREE Full text] [doi: 10.1289/EHP2175] [Medline: 29467105]
27. US Food and Drug Administration. 2019. Trump Administration Combating Epidemic of Youth E-Cigarette Use with Plan to Clear Market of Unauthorized, Non-Tobacco-Flavored E-Cigarette Products URL: https://www.fda.gov/news-events/press-announcements/trump-administration-combating-epidemic-youth-e-cigarette-use-plan-clear-market-unauthorized-non [accessed 2020-01-13]
28. Perrine CG, Pickens CM, Boehmer TK, King BA, Jones CM, de Sisto CL, Lung Injury Response Epidemiology/Surveillance Group. Characteristics of a multistate outbreak of lung injury associated with e-cigarette use, or vaping - United States,

2019. MMWR Morb Mortal Wkly Rep 2019 Oct 4;68(39):860-864 [FREE Full text] [doi: 10.15585/mmwr.mm6839e1] [Medline: 31581168]

29. Lewis N, McCaffrey K, Sage K, Cheng C, Green J, Goldstein L, et al. E-cigarette use, or vaping, practices and characteristics among persons with associated lung injury-Utah, April-October 2019. MMWR Morb Mortal Wkly Rep 2019 Oct 25;68(42):953-956 [FREE Full text] [doi: 10.15585/mmwr.mm6842e1] [Medline: 31647788]
30. Brown CJ, Cheng JM. Electronic cigarettes: product characterisation and design considerations. Tob Control 2014 May;23(Suppl 2):ii4-i10 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051476] [Medline: 24732162]

Abbreviations

DIY: do it yourself
ENDS: electronic nicotine delivery systems
PG: propylene glycol
RBA: rebuildable atomizer
THC: tetrahydrocannabinol
VG: vegetable glycerin

Edited by G Eysenbach; submitted 21.11.19; peer-reviewed by DM Liou, JP Allem, E Da Silva; comments to author 06.01.20; revised version received 25.02.20; accepted 26.02.20; published 02.06.20.

Please cite as:
Massey ZB, Li Y, Holli J, Churchill V, Yang B, Henderson K, Ashley DL, Huang J, Popova L
Modifications to Electronic Nicotine Delivery Systems: Content Analysis of YouTube Videos
J Med Internet Res 2020;22(6):e17104
URL: https://www.jmir.org/2020/6/e17104
doi:10.2196/17104
PMID:32348288

Original Paper

The Use of Mobile Personal Health Records for Hemoglobin A1c Regulation in Patients With Diabetes: Retrospective Observational Study

Dongjin Seo[1*], BS; Yu Rang Park[2*], PhD; Yura Lee[3], MD, PhD; Ji Young Kim[4], MS; Joong-Yeol Park[5], MD, PhD; Jae-Ho Lee[3,6], MD, PhD

[1]Department of Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[3]Department of Information Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[4]Medical Information Office, Asan Medical Center, Seoul, Republic of Korea
[5]Department of Endocrinology and Metabolism, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[6]Department of Emergency Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Jae-Ho Lee, MD, PhD
Department of Information Medicine
Asan Medical Center, University of Ulsan College of Medicine
88 Olympic-ro 43-gil, Songpa-gu
Seoul, 05505
Republic of Korea
Phone: 82 230103350
Email: rufiji@gmail.com

Abstract

Background: The effectiveness of personal health records (PHRs) in diabetes management has already been verified in several clinical trials; however, evidence of their effectiveness in real-world scenarios is also necessary. To provide solid real-world evidence, an analysis that is more accurate than the analyses solely based on patient-generated health data should be conducted.

Objective: This study aimed to conduct a more accurate analysis of the effectiveness of using PHRs within electronic medical records (EMRs). The results of this study will provide precise real-world evidence of PHRs as a feasible diabetes management tool.

Methods: We collected log data of the *sugar* function in the My Chart in My Hand version 2.0 (MCMH 2.0) app from Asan Medical Center (AMC), Seoul, Republic of Korea, between December 2015 and April 2018. The EMR data of MCMH 2.0 users from AMC were collected and integrated with the PHR data. We classified users according to whether they were continuous app users. We analyzed and compared their characteristics, patterns of hemoglobin A_{1c} (HbA_{1c}) levels, and the proportion of successful HbA_{1c} control. The following confounders were adjusted for HbA_{1c} pattern analysis and HbA_{1c} regulation proportion comparison: age, sex, first HbA_{1c} measurement, diabetes complications severity index score, sugar function data generation weeks, HbA_{1c} measurement weeks before MCMH 2.0 start, and generated sugar function data count.

Results: The total number of MCMH 2.0 users was 64,932, with 7453 users having appropriate PHRs and diabetes criteria. The number of continuous and noncontinuous users was 133 and 7320, respectively. Compared with noncontinuous users, continuous users were younger (*P*<.001) and had a higher male proportion (*P*<.001). Furthermore, continuous users had more frequent HbA_{1c} measurements (*P*=.007), shorter HbA_{1c} measurement days (*P*=.04), and a shorter period between the first HbA_{1c} measurement and MCMH 2.0 start (*P*<.001). Diabetes severity–related factors were not statistically significantly different between the two groups. Continuous users had a higher decrease in HbA_{1c} (*P*=.02) and a higher proportion of regulation of HbA_{1c} levels to the target level (*P*=.01). After adjusting the confounders, continuous users had more decline in HbA_{1c} levels than noncontinuous users (*P*=.047). Of the users who had a first HbA_{1c} measurement higher than 6.5% (111 continuous users and 5716 noncontinuous users), continuous users had better regulation of HbA_{1c} levels with regard to the target level, 6.5%, which was statistically significant (*P*=.04).

Conclusions: By integrating and analyzing patient- and clinically generated data, we demonstrated that the continuous use of PHRs improved diabetes management outcomes. In addition, the HbA_{1c} reduction pattern was prominent in the PHR continuous user group. Although the continued use of PHRs has proven to be effective in managing diabetes, further evaluation of its effectiveness for various diseases and a study on PHR adherence are also required.

(J Med Internet Res 2020;22(6):e15372) doi:10.2196/15372

KEYWORDS

personal health record; mobile health; electronic medical record; diabetes mellitus; glycated hemoglobin A

Introduction

Background

Diabetes mellitus is a global issue, and its contribution to numerous complications and increased mortality is well known. Moreover, diabetes prevalence is constantly growing, a trend that might continue until 2030 or longer [1,2]. According to the American Diabetes Association (ADA), diabetes care is mainly based on insulin delivery [3]. According to the Korean Diabetes Association (KDA), the target value of hemoglobin A_{1c} (HbA_{1c}) is recommended to be 6.5% for patients with type 2 diabetes, and antihyperglycemic therapy is mainly considered in Korea. Metformin is considered to be the first-line therapy. However, these traditional drug therapies result in inevitable hypoglycemic events and body weight change. An unachieved glycemic target can only be solved by increasing drugs in mono, dual, or triple therapy [4]. Traditional methods are expensive, and this is becoming a national health care problem [5,6]. To overcome several limitations of traditional diabetes management, mobile health (mHealth) technology and personal health record (PHR) implementation have been suggested as innovative solutions.

In the diabetes management market, new treatments with new devices and apps are being introduced. Most functions of diabetes apps focus on maintaining a blood glucose diary. Some are also connected with blood glucose sensors and treatment devices. Among diabetes apps, *OneTouch Reveal* had the best validation [7]. This app is wirelessly connected to the *OneTouch Verio Flex meter*, making users self-monitor their blood glucose. Blood glucose data are delivered to health care professionals, and users receive text message feedback [8]. Technologies using automatic alarm systems have also been introduced. The Dexcom G6 Continuous Glucose Monitoring system effectively reduced hyperglycemia and also hypoglycemic events with the *Urgent Low Soon* automatic alert system [9]. Monitoring insulin delivery became possible with internet-based connections. *NovoPen 6* and *NovoPen Echo Plus* are called *smart insulin pens*, which can monitor the insulin injection amount and provide both health providers and patients treatment accuracy [10,11].

Previous studies have shown the health improvement of PHR users, thus suggesting that a digital health care system is feasible for improving health behavior and chronic conditions. According to a systematic review, users experienced a positive effect on their health-related behavior and clinical results when using health apps on their mobile devices [12]. Another systematic review in South Korea showed that mHealth interventions were effective in improving self-management behaviors, biomarkers, or patient-reported outcome measures [13]. However, the positive effect of mHealth and PHR interventions is not always ensured.

In diabetes care, PHR and mHealth interventions are expected to be effective treatments. WellDoc, a remote blood glucose monitoring system, was effective in lowering HbA_{1c} levels, thereby improving clinical, behavioral, and diabetes knowledge outcomes [14]. A phone-based treatment and behavioral coaching intervention also improved HbA_{1c} levels [15]. A similar improvement in HbA_{1c} control for type 2 diabetes was seen with another mobile-based intervention [16]. The addition of a tailored mobile coaching system for patients with diabetes showed reduced HbA_{1c} levels and improved diabetes self-management; the results were reproducible and durable [17].

Along with the expectations of the clinical implications of PHRs, some concerns and slightly controversial results have been reported. Despite its advantages, studies have reported the barriers in PHR implementation. Patients are concerned about the security of their health information. Health care providers are concerned about patients altering their own PHR information. Other issues are that there is no practical difference in health outcomes, the use of stand-alone PHRs with electronic medical records (EMRs) and electronic health records, and a low health care literacy rate, which can diminish the benefits of PHRs [18]. Moreover, the barriers associated with patients' age, sex, socioeconomic status, education level, internet and computer access, and health have been reviewed [19]. Contrasting results of the relation between PHR use and diabetes management have been reported. A study using a regression model claimed that there was no association between the increasing number of days of PHR use and better diabetes quality measure profiles [20].

Objectives

In this study, we used a 4-year mobile PHR (mPHR) log and users' EMR data to analyze the effects of diabetes management on the continuous use of the PHR system distributed by a tertiary hospital in South Korea. A study with the earlier version of the mPHR app was conducted to verify characteristics of continuous users [21], and patient-generated health data (PGHD) of continuous users had a higher proportion of a chronic disease diagnosis, such as diabetes, than noncontinuous users [22]. With the new version, we will verify its effect in glycemic control on patients with diabetes. To the best of our knowledge, this is the first study to verify the effectiveness of disease management by integrating a long-term mPHR log and EMR data.

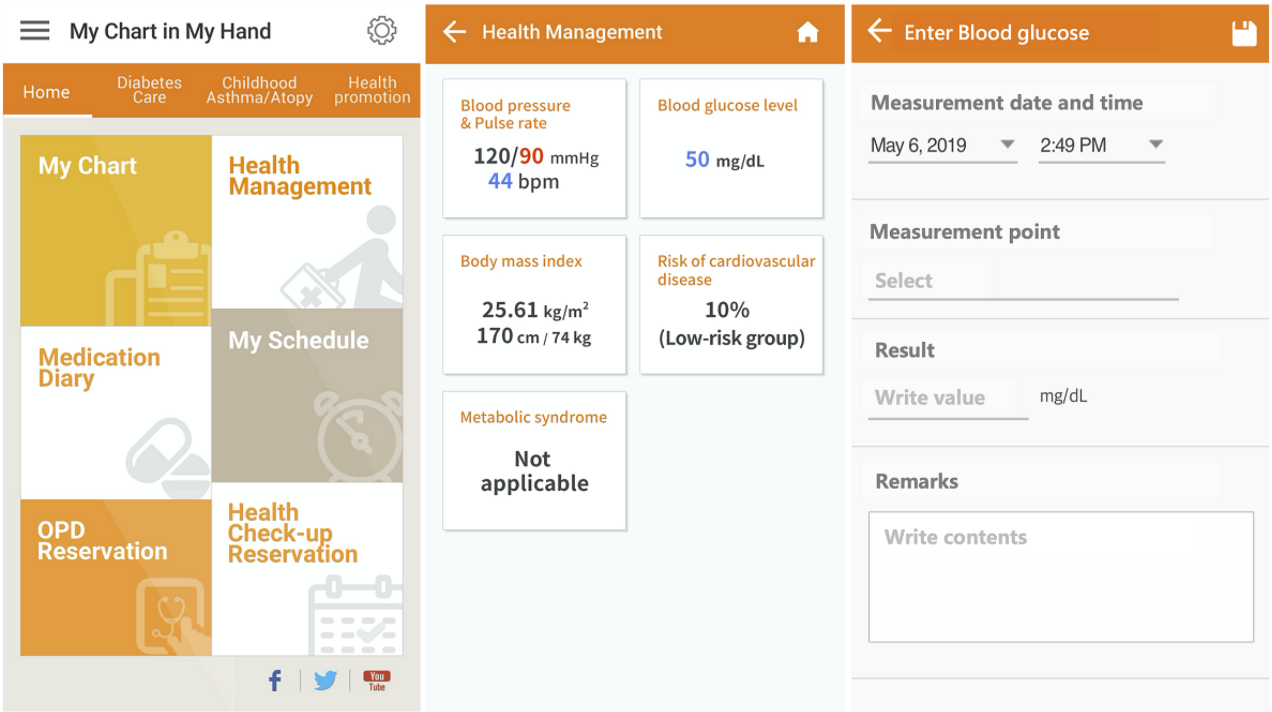
Methods

Data and Mobile Personal Health Record Description

We collected log data from an mPHR app called My Chart in My Hand (MCMH) and their EMR data at the Asan Medical Center (AMC), which is the largest general hospital in South Korea. Launched in January 2011, MCMH is the first mPHR in South Korea; it enables patients to view and manage their own health records [21]. We used the MCMH version 1.0 log to identify patterns of continuous generation of PGHD in specific populations [22]. This study performed a diabetes management analysis using the MCMH version 2.0 log and EMR data. For patients with diabetes, MCMH version 2.0 provides *sugar*, *diabetes calendar*, *insulintreatment*, *food intake*, and *exercise* input functions. Among these functions, we only used the log data of the *sugar* and *diabetes calendar* function; the remaining functions had very few records. The items in Figure 1 show the details of the *sugar* function. Users enter the date, time, situation, and result of their blood glucose measurement in these PGHD functions.

We also gathered demographic and medical record information of patients, such as age, sex, residence, and health information, including hospital visits, HbA_{1c} level, diagnosis, and medication data, using our clinical research data warehouse.

Figure 1. Screenshots of My Chart in My Hand version 2.0. Inputting data in the sugar function follows from the home page to Enter Blood Glucose.



Blood pressure & Pulse rate 120/90 mmHg 44 bpm

Blood glucose level 50 mg/dL

Body mass index 25.61 kg/m² 170 cm / 74 kg

Risk of cardiovascular disease 10% (Low-risk group)

Metabolic syndrome Not applicable

Measurement date and time May 6, 2019 2:49 PM

Result Write value mg/dL

Study Design

MCMH version 2.0 replaced MCMH version 1.0 on December 31, 2015, but some patients had already created their accounts in December 2015 before the replacement. For each user, the records generated in MCMH version 2.0 functions were analyzed, but only records generated after account creation were used.

The user log of the *sugar* function contained user access ID and time stamps of data input. We gathered the HbA_{1c} measurement results of MCMH version 2.0 users from January 2014 to November 2018.

For user selection, we used the criteria of diabetes for diagnosis. First, the criterion of Glasheen et al [23] was adopted: a user should have one or more International Classification of Diseases 10th Revision (ICD-10) diabetes codes in the diagnosis record, which are E08, E09, E10, E11, and E13. Second, the HbA_{1c} cutoff value of 6.5% for diagnosing diabetes was used [24]. For the complication classification and diabetes complications severity index (DCSI) scoring, the selected complication fields from the diagnosis record were retinopathy, nephropathy, neuropathy, cerebrovascular, cardiovascular, peripheral vascular disease, and metabolic complications. DCSI scoring used the criteria of the study by Glasheen et al [23]. However, urine laboratory data were not included in DCSI scoring because of its unavailability. Above all, we classified all diseases according to ICD-10.

The criterion for whether a user was a continuous user was adopted from the PGHD pattern analysis study of MCMH version 1.0: a user entering data in the *sugar* function at least once per week and doing so for at least four weeks (28 days) [22].

We analyzed the pattern of HbA_{1c} levels with the trend line slope of HbA_{1c} levels. The fluctuation of HbA_{1c} levels was compared with the *r*-squared value of the trend line and the standard deviation of the patient's HbA_{1c} level.

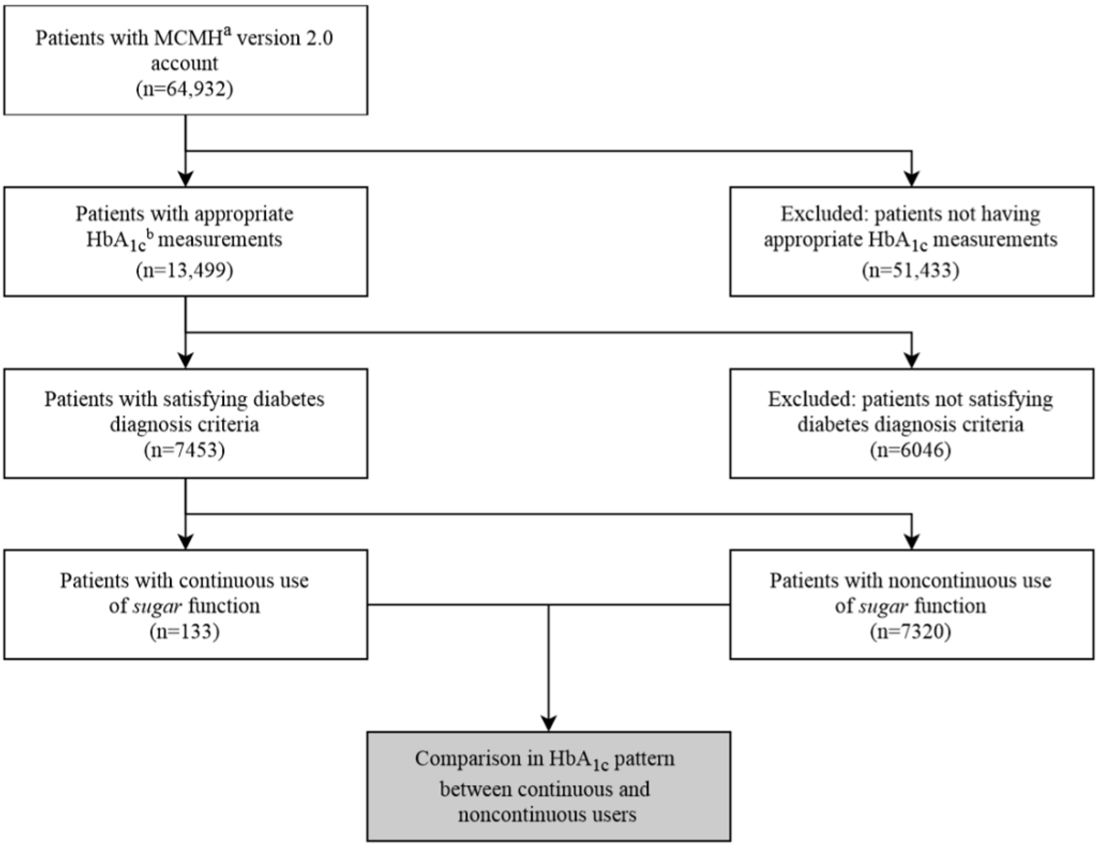
In this study, the trend line slope considerably depended on the measurement days between the first and last HbA_{1c} measurement. Therefore, we created a patient filter called *appropriate* HbA_{1c} *measurement*. This criterion excluded patients with short periods between measures because a short period will lead to an exaggeratedly steep slope, which is inappropriate for the analysis. The criterion for an appropriate HbA_{1c} measurement is patients should have at least two HbA_{1c} measurements and the period between the first and last HbA_{1c} measurement should be over 100 days. To normalize the effect of measurement days between the first and last HbA_{1c} measurement, we defined a variable called *decline*. *Decline* is defined as a trend line slope times the period (in days) divided by 100. This normalization is represented in the equation in Multimedia Appendix 1.

This study was approved by the Institutional Review Board (IRB) of the AMC (IRB number: 2018-0321). The need for informed consent was waived by the ethics committee because this study utilized routinely collected log data that were anonymously managed at all stages, including during data cleaning and statistical analyses.

Study Participants

Figure 2 shows the patient selection flow in this study. Among 64,932 users who downloaded and created an MCMH version 2.0 account, we first excluded 51,433 users with inappropriate HbA_{1c} measurements. We considered 13,499 users with the appropriate HbA_{1c} measurements, excluded 6046 users without diabetes, and selected 7453 users with diabetes.

Figure 2. Patient inclusion and exclusion criteria (white boxes) and flow through the study. The gray box shows user hemoglobin A_{1c} (HbA_{1c}) analyses. Criteria for appropriate HbA_{1c} measurement: two or more HbA_{1c} measurements, duration of the first and last measurement over 100 days, and creating My Chart in My Hand version 2.0 account during HbA_{1c} measurement. Criteria for diabetes diagnosis: having International Classification of Diseases 10th Revision code E08, E09, E10, E11, or E13 or first HbA_{1c} measurement ≥6.5%. Criteria for continuous use of sugar function: patient-generated health data entered in the sugar function at least once per week and used for at least 28 days. [a]HbA_{1c}: hemoglobin A_{1c}; [b]MCMH: My Chart in My Hand.



Data Analysis

We first compared the general characteristics of continuous (n=133) and noncontinuous users (n=7320). The following characteristics were compared: age, gender proportion, *sugar* and *diabetes calendar* function use pattern, HbA_{1c} measurement pattern, HbA_{1c} value, DCSI score, and complication proportion. A Student *t* test was conducted for the comparison of age, the number of HbA_{1c} measurements, measurement days, and measurement days before MCMH version 2.0 start. A Wilcoxon rank-sum test was used for individual *sugar* and *diabetes calendar* function data generation, HbA_{1c} measure frequency, first HbA_{1c} measurement, and DCSI score comparison. The median test was used for the individual *sugar* and *diabetescalendar* function data generation comparison. The Z test was conducted for *sugar* and *diabetes* function generation user proportion, first HbA_{1c} measurement over 6.5% proportion, and complications proportion comparisons. For gender proportion comparison and DCSI score distribution, a chi-square test was used.

Next, comparative analyses of HbA_{1c}*decline*, *r*-squared value, and standard deviation between continuous and noncontinuous users were performed. We used the Shapiro-Wilk test and D'Agostino K-squared test to determine if these data followed a normal distribution. HbA_{1c}*decline*, *r*-squared value, and standard deviation were compared using the Wilcoxon rank-sum test. For confounder adjustment, we used an analysis of covariance (ANCOVA) with some variables: continuous use, age, sex, first HbA_{1c} measurement, DCSI, *sugar* function data generation weeks, HbA_{1c} measurement in weeks before MCMH version 2.0 start, and *sugar* function data generation count.

Finally, the Z test was conducted for comparing the proportions of 4 groups between continuous and noncontinuous users. The 4 groups were divided by whether the first HbA_{1c} measurement was higher or lower than 6.5% and whether the last HbA_{1c} measurement was higher or lower than 6.5%. For confounder adjustment, multivariable logistic regression was used for users with the first HbA_{1c} measurement over 6.5%. The same variables, as used in ANCOVA, were used for logistic regression. Data analyses were conducted using *Python* 3.6.7, with *Jupyter Notebook*.

Results

Overall Characteristics

Within 29 months of operation of MCMH version 2.0, 64,932 users created an account and logged in at least once. Among these users, 7453 users were selected on the basis of the inclusion criteria of this study. Approximately 1.78% (133/7453) of these users were continuous users, and 98.22% (7320/7453) were noncontinuous users. Continuous and noncontinuous users had no statistically significant difference in the number of HbA_{1c} measurements and the period between the first and last HbA_{1c} measurements.

Table 1 summarizes the results of a basic characteristic analysis between continuous and noncontinuous users. In Table 1, measure frequency refers to the number of measurements per day, measurement days refers to days between the first and last HbA_{1c} measurement, and measurement days before MCMH version 2.0 start refers to days between the first HbA_{1c} measurement and MCMH version 2.0 account generation period. Compared with noncontinuous users, continuous users were younger (mean 53.59, SD 9.89 years vs mean 57.58, SD 11.95 years, respectively) and had a higher male proportion (110/133, 82.7% vs 4859/7320, 66.38%, respectively), which was statistically significant (both *P*<.001). The number of HbA_{1c} measurements was not significantly different. The frequency and period between the first and last measurements exhibited a significant difference between continuous and noncontinuous users (*P*=.007 and *P*=.04, respectively). The proportion of patients with the first HbA_{1c} measurement below 6.5% had no significant difference (*P*=.14), but continuous users had a higher first HbA_{1c} measurement, and this was statistically significant (*P*=.01). Furthermore, among continuous users, there were a higher proportion of users who generated data in the sugar function and diabetes calendar function (both *P*<.001). Continuous users also entered more sugar and diabetes calendar data (both *P*<.001). The DCSI score had no significant difference (*P*=.99). The proportion of complications, defined by the DCSI criteria, also showed no significant difference between continuous and noncontinuous users. Although the difference was statistically insignificant, retinopathy and cardiovascular complications had a proportional difference.

The DCSI score proportion of continuous and noncontinuous users had no significant difference in the chi-square test. This can be found in Multimedia Appendix 2. Among the 14 DCSI scores, those with zero proportion in both patient groups (scores 10, 12, and 13) were excluded in the analysis using the chi-square test, because calculation with the chi-square test is only possible when each score does not have zero proportion in any group.

Table 1. General characteristics of continuous and noncontinuous users.

Variables	Users		Total (N=7453)	*P* value[a]
	Continuous (n=133)	Noncontinuous (n=7320)		
Age (years), mean (SD)	53.59 (9.89)	57.58 (11.95)	57.51 (11.92)	<.001
Sex, n (%)				<.001
Male	110 (82.7)	4859 (66.37)	4969 (66.67)	
Female	23 (17.3)	2461 (33.62)	2484 (33.33)	
Sugar function				
Data generated by users, n (%)	133 (100.0)	289 (3.95)	422 (5.66)	<.001
Total data generated, n	22,350	1345	23,695	—[b]
Individually generated data				<.001
Mean (SD)	168.0 (204.0)	0.2 (1.8)	3.2 (35.1)	
Median (IQR)	97 (43-186)	0 (0-0)	0 (0-0)	
Diabetes calendar function				
Data generated by users, n (%)	133 (100.0)	297 (4.06)	430 (5.77)	<.001
Total data generated, n	16,407	1453	17,860	—
Individually generated data				<.001
Mean (SD)	123.4 (143.3)	0.2 (4.0)	2.4 (25.4)	
Median (IQR)	67 (35-145)	0 (0-0)	0 (0-0)	
HbA_{1c}[c], mean (SD)				
Number of measurements	12.44 (6.90)	11.90 (6.82)	11.92 (6.82)	.38
Measure frequency	0.011 (0.010)	0.009 (0.005)	0.009 (0.005)	.007
Measurement days	1254 (461)	1336 (445)	1335 (446)	.04
Measurement days before MCMH[d] version 2.0 start	546 (348)	712 (377)	710 (377)	<.001
First HbA_{1c} measurement ≥6.5%, n (%)	111 (83.4)	5716 (78.09)	5827 (78.18)	.14
First HbA_{1c} measurement, mean (SD)	7.86 (1.78)	7.51 (1.62)	7.51 (1.62)	.01
DCSI[e], mean (SD)	1.17 (1.65)	1.15 (1.64)	1.15 (1.64)	.99
Complications, n (%)				
Retinopathy or ophthalmic	31 (23.3)	1516 (20.71)	1547 (20.75)	.46
Nephropathy	13 (9.8)	765 (10.45)	778 (10.44)	.80
Neuropathy	23 (17.3)	1267 (17.31)	1290 (17.31)	>.99
Cerebrovascular	20 (15.0)	950 (13.00)	970 (13.01)	.48
Cardiovascular	16 (12.0)	1366 (18.7)	1382 (18.54)	.05
Peripheral vascular disease	1 (0.8)	59 (0.8)	60 (0.81)	.94
Metabolic complications	1 (0.8)	37 (0.5)	38 (0.51)	.69

[a] Chi-square test or Z test (for categorical variables); Student *t* test or Wilcoxon rank-sum test (for continuous variables).

[b] Statistical comparison was not conducted in total generated data of sugar and diabetes calendar function.

[c] HbA_{1c}: hemoglobin A_{1c}.

[d] MCMH: My Chart in My Hand.

[e] DCSI: diabetes complications severity index.

Hemoglobin A_{1c} Pattern Analysis According to Continuous Use

Figure 3 shows the trend of the HbA_{1c} pattern for continuous and noncontinuous users. The HbA_{1c} *decline* of continuous and noncontinuous users was also compared. The HbA_{1c} *decline* (mean −0.00533, SD 0.0144) in continuous users was significantly steeper than that of noncontinuous users (mean −0.00278, SD 0.0137; *P*=.02). The SD of continuous users (mean 0.832, SD 0.574) was significantly higher than that of noncontinuous users (mean 0.719, SD 0.541; *P*=.005). However, the *r*-squared value had no statistically significant difference between continuous and noncontinuous users (*P*=.40).

When adjusting confounders that can contribute to the *decline*, continuous use had a statistically significant effect (*P*=.047) on making *decline* steeper, as seen in Table 2. In addition, age, first HbA_{1c} measurement, DCSI, weeks of *sugar* function data generation, and HbA_{1c} measurement in weeks before MCMH version 2.0 start showed statistically significant effects (*P*=.004; *P*<.001; *P*=.01; *P*=.003; *P*<.001, respectively).

Figure 3. Hemoglobin A_{1c} (HbA_{1c}) patterns (decline, r-squared value, and SD) of continuous and noncontinuous users. The x-axis is the percentage of days past from the first HbA_{1c} measurement compared with the period between the first and last HbA_{1c} measurements. The dashed lines are the HbA_{1c} decline of each patient. The slope and y-axis intercept of the continuous lines indicates the mean of slope and y-axis of patients, respectively.

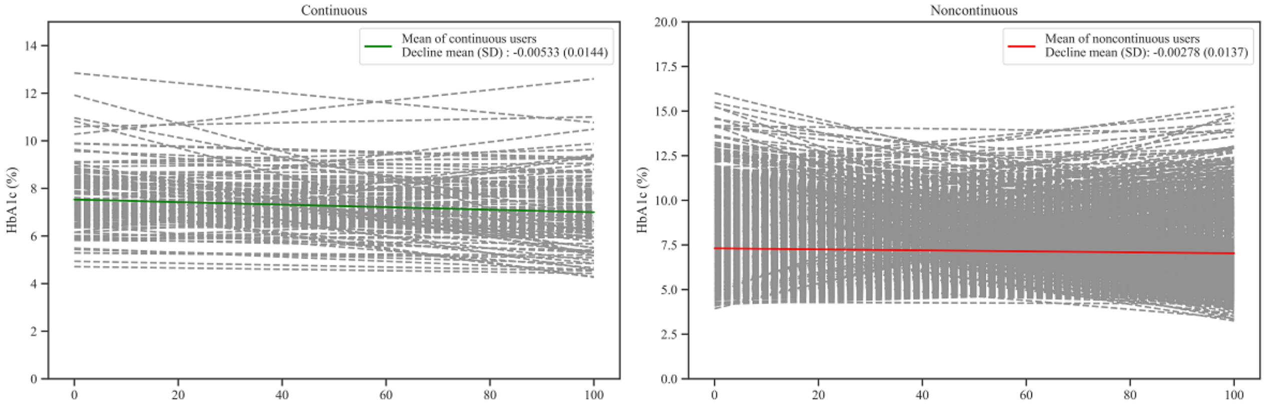


Table 2. Results of adjusting confounders with the analysis of covariance in decline comparison.

Variables	*F* test (*df*=1)	*P* value
Continuous users	3.94	.047
Age (years)	8.07	.004
Sex	0.17	.68
First HbA_{1c}[a] measurement	3054.90	<.001
DCSI[b]	6.45	.01
Sugar function data generation (weeks)	8.68	.003
HbA_{1c} measurement weeks before MCMH version 2.0 start	154.25	<.001
Generated *sugar* function data count	0.03	.86

[a] HbA_{1c}: hemoglobin A_{1c}.

[b] DCSI: diabetes complications severity index.

Comparison of Hemoglobin A_{1c} Regulation With Target Level in Continuous Use

Table 3 lists the proportion with regard to HbA_{1c} patterns. The proportion of users with the first HbA_{1c} measurement higher than 6.5% and the last HbA_{1c} measurement lower than 6.5% had a statistical difference (*P*=.01). Among users with the first HbA_{1c} measurement lower than 6.5%, the proportion of patients with the last HbA_{1c} measurement lower than 6.5% and the last HbA_{1c} measurement higher than 6.5% had no significant difference (*P*=.34 and *P*=.29, respectively). No significant difference was found between proportions of patients with the first HbA_{1c} measurement of 6.5% or higher and the last HbA_{1c} measurement higher than 6.5% (*P*=.41).

Similar to the *decline* analysis, the result of confounder adjustment by logistic regression for users with a high first HbA_{1c} measurement is summarized in Table 4. The continuous use of MCMH version 2.0 had a statistically significant effect in helping users move from an HbA_{1c} measurement above 6.5% to an HbA_{1c} measurement below 6.5% (*P*=.04). In addition, age, first HbA_{1c} measurement, and HbA_{1c} measurement in weeks before MCMH version 2.0 start showed statistically significant effects (all: *P*<.001).

Table 3. Pre– and post–hemoglobin A_{1c} management comparison by continuous use.

HbA_{1c}[a] pattern	Users		*P* value
	Continuous (n=133)	Noncontinuous (n=7320)	
First measurement <6.5%			
Last measurement			
<6.5%, n (%)	15 (11.3)	1040 (14.21)	.34
≥6.5%, n (%)	7 (5.3)	564 (7.70)	.29
First measurement ≥ 6.5%			
Last measurement			
<6.5%, n (%)	38 (28.6)	564 (7.70)	.01
≥6.5%, n (%)	73 (54.9)	4278 (58.44)	.41

[a] HbA_{1c}: hemoglobin A_{1c}.

Table 4. The result of logistic regression against users with a high first hemoglobin A_{1c} measurement (n=111 continuous and n=5716 noncontinuous users).

Variables	Coefficient	*P* value
Constant	1.640	<.001
Continuous	0.618	.04
Age (years)	−0.010	<.001
Sex	−0.085	.20
First HbA_{1c}[a] measurement	−0.171	<.001
DCSI[b]	−0.041	.05
Sugar function data generation (weeks)	−0.004	.23
HbA_{1c} measurement in weeks before MCMH[c] version 2.0 use start	−0.008	<.001
Generated *sugar* function data count	−0.001	.52

[a] HbA_{1c}: hemoglobin A_{1c}.

[b] DCSI: diabetes complications severity index.

[c] MCMH: My Chart in My Hand.

Discussion

Principal Findings

For the following reasons, this study supports the use of mPHRs as an effective platform for diabetes management by integrating patient-generated health and clinical data from PHRs and EMRs, respectively. First, analyzing the characteristics of continuous users of MCMH version 2.0, male patients with a high HbA_{1c} level seemed to use MCMH version 2.0 more continuously. Second, the continuous use of PHRs resulted in a higher decrease of HbA_{1c} levels and enhanced the regulation of high HbA_{1c} levels of patients to the target range. Therefore, male users with high HbA_{1c} levels had a higher decrease in HbA_{1c} levels and improved HbA_{1c} regulation to the target level. By analyzing the characteristics of continuous users and their HbA_{1c} patterns, we also suggest the use of mPHR as a diabetes care support tool enabling personalized management.

This study is unique when compared with previous studies on the basis of the following characteristics. First, we suggested the health improvement effect of mPHRs on the basis of the integration of PHRs and EMRs. In this study, we expected two benefits of integrating PHRs and EMRs. One is suggesting a different methodology for real-world data analysis and presenting additional real-world evidence, which supports previous studies. Another is ensuring a high-quality data analysis is conducted. There are many previous studies implying the advantages of PHRs and PGHD with positive conclusions of the use of mPHRs [14-17]. The results of these studies were collected on the basis of clinical trials such as nonblinded, open-label randomized controlled trials (RCTs) and cluster-randomized trial designs. As a real-world data analysis covers bias limitations in RCTs and can handle unknown factors of PHRs, the results of a real-world data analysis provide strong and necessary support to previous RCTs [25]. Moreover, the integration of EMRs gave high-quality HbA_{1c} data and diagnosis data, which made the analysis more precise.

Second, previous studies mainly discussed about the decrease in HbA_{1c} levels as an advantage of using PHRs. However, as the main goal of glycemic control is regulating a patient's HbA_{1c} level to the recommended range, we compared both HbA_{1c} decrease and proportions of patients who initially had a high HbA_{1c} level but their HbA_{1c} level decreased to a low value. According to the 2015 and 2019 diabetes management guidelines from the KDA, the recommended target HbA_{1c} level is 6.5% in patients with type 2 diabetes, and this differs from the guideline by the ADA [4,26,27]. As this study was conducted in AMC, South Korea, we used the guidelines from KDA and defined the cutoff value of the HbA_{1c} level as 6.5%. Recent studies recommend that patients with severe diabetes mellitus should be controlled to lower than 7%, depending on the severity and complications of diabetes [28-30]. Moreover, a stable decrease in blood glucose levels is also an important task in glycemic control. We also focused on the *r*-squared value of the trend line and SD as an indicator of stabilized HbA_{1c} decrease, but we could not achieve any outstanding results.

Overall User Characteristics

Analyzing users who had access to MCMH version 1.0 indicated that these users visited hospitals more with chronic diseases [21]. Continuous users were younger than noncontinuous users ($P<.001$), and there was a significant difference in sex proportion; the continuous user group had a higher male ratio ($P<.001$). In previous research, groups that used a PHR system had young users and a high proportion of males or generated more PGHD, especially those related to diabetes [21,22]. This is because male users aged between 51 and 70 years tend to adopt the PHR system [31]. In addition, in this study, the HbA_{1c} level in continuous users was measured for a shorter period ($P=.04$) and more frequently ($P=.007$) than noncontinuous users. However, the number of HbA_{1c} measurements had no significant difference between continuous and noncontinuous user groups. In South Korea, the social health insurance program was introduced with the 1977 National Health Insurance Act. This program was thereafter progressively rolled out to the general public, and it finally achieved universal coverage in 1989. According to the National Health Insurance Act, the criteria for the method, procedure, scope, and upper limit of health care shall be prescribed by the Ministry of Health and Welfare [17].

National insurance only supports up to 6 HbA_{1c} tests per year, in accordance with the National Health Insurance Act. First, we considered the number of HbA_{1c} measurements as another indicator of diabetes severity. This is because well-controlled patients typically undergo HbA_{1c} tests twice a year, whereas poorly controlled individuals undergo testing 4 times a year [32]. However, the number of measurements seems to be similar because of the policy in South Korea. Although continuous users had shorter periods (approximately 80 days) between the first and last measurements, this group took HbA_{1c} tests more frequently. This may be because of the increase in hospital visits, along with more satisfaction and loyalty to the hospital [33]. To compare diabetes severity, the proportion of patients with an HbA_{1c} level of 6.5% or above, a first HbA_{1c} level measurement, and a DCSI score distribution were compared between continuous and noncontinuous groups. The two groups had no significant difference in the proportion of high HbA_{1c} levels and DCSI distribution; however, continuous users had a higher HbA_{1c} level ($P=.01$). Retinopathy patients tended to use MCMH version 2.0 more continuously, but the complication proportion also had an insignificant difference between the two groups. Except for the first HbA_{1c} level measurement, most diabetic-related baseline characteristics appeared to have no significant difference, and the first HbA_{1c} measurement can be adjusted as confounders in an additional analysis. By using PHR and EMR integration, the general characteristics and severity of diabetes were compared.

As the period of HbA_{1c} measurement before MCMH version 2.0 use was shorter in the continuous group ($P<.001$), continuous users seemed to have an earlier MCMH version 2.0 start compared with noncontinuous users. In addition, continuous users tended to use the *sugar* and *diabetes calendar* functions more and generate more data. This was because continuous users tended to use MCMH version 2.0 functions with fewer burdens.

Verifying the Effect of Personal Health Record Use in Hemoglobin A_{1c} Control

The main advantage of PHRs and PGHD is health improvement, especially in diabetes. Among the types of diabetes management, determining the change in HbA_{1c} levels was the most effective method to verify the effectiveness of PHRs in the real world. The results of this study indicate that continuous users had a larger *decline*; a greater increase in HbA_{1c} levels was observed in users who continuously used the diabetes management–related *sugar* function in MCMH version 2.0. As *decline* is the result of the trend line slope normalized to 100 days, the value itself also refers to the change in the HbA_{1c} level. For example, HbA_{1c} was 6.9% on January 1, 2014, and HbA_{1c} was 6.4% on October 19, 2018, in one particular continuous user; therefore, the decline value was −0.0044, which means that this patient's change in HbA_{1c} level was approximately −0.44% (100 times the value of decline). Thus, the decrease in HbA_{1c} levels in continuous users was approximately 1.9 times that in noncontinuous users. The result of ANCOVA shows that along with continuous use, other factors were also important: age, first HbA_{1c} measurement, DCSI, duration of using the *sugar* function, and HbA_{1c} measurement period before using MCMH version 2.0. Glycemic control is important for reducing both microvascular risk and emergent risk for myocardial infarction and death [34]. This indicates that the group that continuously used PHRs had health improvement with a decreasing trend of HbA_{1c} levels.

In glycemic control, it is important to reduce not only blood glucose levels but also hypoglycemic events [35]. Traditional diabetes care includes insulin delivery using syringes, pens, or pumps [3]. Although hypoglycemic side effects can occur with multiple daily injections and continuous subcutaneous insulin injection, the invasive characteristic of such forms of care is an inevitable disadvantage [36-39]. In this study, we tried to minimize the risk of hypoglycemic events in PHR-implemented diabetes management by using stability indicators, *r*-squared value and *SD*. However, stability was not ensured. In fact, a previous study showed increased glucose stability with the use of an internet-based glucose monitoring system [40]. This indicates that patients can improve hyperglycemia and hypoglycemia management by using PHRs with a blood glucose meter through continuous glucose monitoring diabetic care.

The goal of decreasing the HbA_{1c} level is to prevent the occurrence and aggravation of diabetic complications. Although the criterion for HbA_{1c} in a diagnostic test for diabetes has been recommended by the American Association of Clinical Endocrinologists and ADA, it is an "acceptable complementary diagnostic test for diabetes in Korean patients" [28,41]. Among the many glycemic controls, the tight regulation of HbA_{1c} levels is essential for health improvement and for lowering complication risks such as diabetic retinopathy [42]. In addition, the tight glycemic control of HbA_{1c} levels to 7.0% induces a lower risk of fracture in elderly patients with diabetes [43]. When comparing the ratio of patients with HbA_{1c} levels above and below 6.5% before and after the use of MCMH version 2.0, the group that continuously used MCMH version 2.0 had a higher proportion of regulated patients; initially, the first HbA_{1c} level measurement was over 6.5%, and then it reduced to lower than 6.5%. In addition, among users with the first HbA_{1c} level measurement over 6.5%, the logistic regression results showed that regulation was associated not only with continuous use but also with age, first HbA_{1c} level measurement, and how fast MCMH version 2.0 was adapted. The data generation amount was thought to be important too, but it was statistically insignificant. Therefore, we can claim that the improvement of HbA_{1c} levels by PHR use can eventually affect diabetes management by controlling HbA_{1c} levels to 6.5% in practice.

Limitations of This Research

The main limitation of this study is the concern of general biases in real-world studies: selection bias, information bias, recall bias, and detection bias [44]. As this study mainly focused on analyzing real-world data, strict criteria and inevitable exclusion are necessary, leading to concerns in selection bias and detection bias. However, the criteria for the comparison group were the same, and despite including and excluding many patient criteria and comparing with the MCMH 1.0 user analysis, the study scale is almost similar [22]. The size of the continuous user groups is sometimes larger than that used in other RCT studies and had little baseline differences in diabetic severity [17]. As MCMH version 2.0 data are PGHD, continuous use can only be analyzed by its log data, which does not represent adherence to the app and can lead to information bias. On the contrary, we note that information bias that can occur in HbA_{1c} level scaling can be controlled with the integration of EMRs. This integration helped in reducing recall bias in diabetes and complication diagnosis.

Time scale is also another limitation. In RCTs, the HbA_{1c} measurement point, the app account creation point, and app use frequency can be controlled and optimized for convenient data analysis. However, in real-world data, patients have diverse points of HbA_{1c} measurement and MCMH version 2.0 starting points. Even though there were limitations with regard to missing data, inappropriate data, and ambiguous time scale standards, we used patient selection criteria to choose patients who can be analyzed and used the *decline* factor to monitor the HbA_{1c} level for minimizing the effect of irregular time points. The *decline* factor is a variable that has been coined for the purpose of this study and has an uncertain clinical rationale. However, as the *decline* variable also implies a decrease in HbA_{1c} levels, and the decreasing trend is being maintained, the quantitative comparison of *decline* between groups is meaningful. In diabetes care, lowering HbA_{1c} levels to the target level and maintaining the decreased HbA_{1c} level is the primary goal. Thus, the *decline* is a reasonable variable for analysis in studies with data having unspecific HbA_{1c} measurement points.

An additional limitation is that AMC is a territorial hospital, and almost all the study patients are residing in South Korea. The small size of the study population and short duration are other limitations. The low frequency of PHR data generation and short-term MCMH version 2.0 operation is not an ideal database for analyzing chronic diseases such as diabetes. A larger study size and longer study duration will provide stronger real-world evidence of the clinical meaning of PHRs.

On the basis of the proportion of continuous and noncontinuous users, further research for encouraging patients to use PHRs more continuously is essential. In this study, continuous users had better diabetes management outcomes than noncontinuous users. However, continuous users were only 1.78% (133/7453) of the study population and were only 0.20% (133/64,932) of users who started using MCMH version 2.0. Thus, studies for maintaining active PGHD-generating users and turning noncontinuous users into continuous users are necessary. Finding out whether giving health-related advice on the basis of MCMH version 2.0 encourages patients to use a PHR app for changing app use patterns needs to be studied to prevent usability issues [45]. Furthermore, for personalized PHR advice, if larger and better quality of data is provided, the glycemic control outcome analysis by treatment is important. Further studies in diverse territories and a deeper analysis of MCMH version 2.0 should be performed to prove the effectiveness of PHRs as a diabetes management tool in decreasing HbA_{1c} levels.

Conclusions

By integrating and analyzing patient- and clinically generated data, the continuous use of PHRs improves diabetes management outcomes. A greater decrease in HbA_{1c} levels was observed in continuous users, and HbA_{1c} levels were regulated to the target level in continuous users compared with noncontinuous users. Previous clinical trials and the results of this study proved that PHRs are effective in managing diabetes. However, further evaluation of the effectiveness of PHRs in various diseases and studies for adherence to PHRs are needed. A larger study population and longer duration will be necessary for the accurate analysis of the clinical rationale of PHRs on chronic diseases.

Acknowledgments

The authors would like to thank the Medical Information Office of AMC for providing log data of the mobile EMR and supporting data analysis and interpretation. This study was supported by a grant of the Research and Development Project, Ministry of Trade, Industry and Energy, Republic of Korea (no. 20004503) and a grant of the National Research Foundation of Korea funded by the Korean government (Ministry of Science and ICT; no NRF-2019M3E5D4064682).

Authors' Contributions

DS, YP, and JL conceived and designed the study; DS, YL, and JK reviewed records and collected the data; DS analyzed the data; DS and YP wrote the manuscript; and YP, JP, and JL reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Formula of decline.
[PNG File , 5 KB - jmir_v22i6e15372_app1.png]

Multimedia Appendix 2
Diabetes complications severity index score proportion comparison of continuous and noncontinuous users.
[PNG File , 26 KB - jmir_v22i6e15372_app2.png]

References

1. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract 2010 Jan;87(1):4-14. [doi: 10.1016/j.diabres.2009.10.007] [Medline: 19896746]
2. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004 May;27(5):1047-1053. [doi: 10.2337/diacare.27.5.1047] [Medline: 15111519]
3. American Diabetes Association. 7. Diabetes Technology: Standards of Medical Care in Diabetes-2019. Diabetes Care 2019 Jan;42(Suppl 1):S71-S80. [doi: 10.2337/dc19-S007] [Medline: 30559233]
4. Kim MK, Ko S, Kim B, Kang ES, Noh J, Kim S, Committee of Clinical Practice Guidelines, Korean Diabetes Association. 2019 clinical practice guidelines for type 2 diabetes mellitus in Korea. Diabetes Metab J 2019 Aug;43(4):398-406 [FREE Full text] [doi: 10.4093/dmj.2019.0137] [Medline: 31441247]
5. Riddle MC, Herman WH. The cost of diabetes care-an elephant in the room. Diabetes Care 2018 May;41(5):929-932. [doi: 10.2337/dci18-0012] [Medline: 29678864]
6. American Diabetes Association. Economic costs of diabetes in the US in 2017. Diabetes Care 2018 May;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]
7. Rodríguez AQ, Wägner AM. Mobile phone applications for diabetes management: A systematic review. Endocrinol Diabetes Nutr 2019 May;66(5):330-337. [doi: 10.1016/j.endinu.2018.11.005] [Medline: 30745121]
8. Grady M, Katz LB, Cameron H, Levy BL. Diabetes app-related text messages from health care professionals in conjunction with a new wireless glucose meter with a color range indicator improves glycemic control in patients with type 1 and type 2 diabetes: randomized controlled trial. JMIR Diabetes 2017 Aug 7;2(2):e19 [FREE Full text] [doi: 10.2196/diabetes.7454] [Medline: 30291092]
9. Derdzinski M, Welsh J, Puhr S, Walker TC, Parker A, Jimenez A. 391-P: Hypoglycemia Reductions with the Dexcom G6 CGM System's Predictive Alert. Diabetes 2019 Jun;68(Supplement 1):391-P. [doi: 10.2337/db19-391-p]
10. Adolfsson P, Hartvig NV, Kaas A, NygÅRd NK, MÅRdby AC, Hellman J. 1076-P: Increased Time-in-Range (TIR) Observed after Introduction of a Connected Insulin Pen. Diabetes 2019;68(Supplement 1):1076-P. [doi: 10.2337/db19-1076-p]
11. Adolfsson P, Hartvig NV, Kaas A, Knudsen NN, MÅRdby A, MøLler JB, et al. 126-LB: Improved Insulin Adherence after Introduction of a Smart Connected Insulin Pen. Diabetes 2019 Jun;68(Supplement 1):126-LB. [doi: 10.2337/db19-126-lb]
12. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
13. Yi JY, Kim Y, Cho Y, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
14. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]

15. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
16. Yu Y, Yan Q, Li H, Li H, Wang L, Wang H, et al. Effects of mobile phone application combined with or without self-monitoring of blood glucose on glycemic control in patients with diabetes: A randomized controlled trial. J Diabetes Investig 2019 Sep;10(5):1365-1371 [FREE Full text] [doi: 10.1111/jdi.13031] [Medline: 30815973]
17. Lee DY, Park J, Choi D, Ahn H, Park S, Park C. The effectiveness, reproducibility, and durability of tailored mobile coaching on diabetes management in policyholders: A randomized, controlled, open-label study. Sci Rep 2018 Feb 26;8(1):3642 [FREE Full text] [doi: 10.1038/s41598-018-22034-0] [Medline: 29483559]
18. Lester M, Boateng S, Studeny J, Coustasse A. Personal health records: beneficial or burdensome for patients and healthcare providers? Perspect Health Inf Manag 2016;13:1h [FREE Full text] [Medline: 27134613]
19. Showell C. Barriers to the use of personal health records by patients: a structured review. PeerJ 2017;5:e3268 [FREE Full text] [doi: 10.7717/peerj.3268] [Medline: 28462058]
20. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
21. Lee G, Park JY, Shin S, Hwang JS, Ryu HJ, Lee JH, et al. Which users should be the focus of mobile personal health records? Analysis of user characteristics influencing usage of a tethered mobile personal health record. Telemed J E Health 2016 May;22(5):419-428 [FREE Full text] [doi: 10.1089/tmj.2015.0137] [Medline: 26447775]
22. Park YR, Lee Y, Kim JY, Kim J, Kim HR, Kim Y, et al. Managing patient-generated health data through mobile personal health records: analysis of usage data. JMIR Mhealth Uhealth 2018 Apr 9;6(4):e89 [FREE Full text] [doi: 10.2196/mhealth.9620] [Medline: 29631989]
23. Glasheen WP, Renda A, Dong Y. Diabetes Complications Severity Index (DCSI)-Update and ICD-10 translation. J Diabetes Complications 2017 Jun;31(6):1007-1013 [FREE Full text] [doi: 10.1016/j.jdiacomp.2017.02.018] [Medline: 28416120]
24. World Health Organization. Geneva: World Health Organization; 2011. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation URL: https://www.who.int/diabetes/publications/report-hba1c_2011.pdf [accessed 2020-03-12]
25. Jadad AR, Enkin MW. Bias in randomized controlled trials. In: Jadad AR, Enkin MW, editors. Randomized Controlled Trials: Questions, Answers and Musings. Paris: BMJ Books; 2007:29-47.
26. Ha KH, Kim DJ. Current status of managing diabetes mellitus in Korea. Korean J Intern Med 2016 Sep;31(5):845-850 [FREE Full text] [doi: 10.3904/kjim.2016.253] [Medline: 27604796]
27. Fruh SM. Obesity: Risk factors, complications, and strategies for sustainable long-term weight management. J Am Assoc Nurse Pract 2017 Oct;29(S1):S3-14 [FREE Full text] [doi: 10.1002/2327-6924.12510] [Medline: 29024553]
28. Handelsman Y, Bloomgarden ZT, Grunberger G, Umpierrez G, Zimmerman RS, Bailey TS, et al. American association of clinical endocrinologists and american college of endocrinology - clinical practice guidelines for developing a diabetes mellitus comprehensive care plan - 2015. Endocr Pract 2015 Apr;21(Suppl 1):1-87 [FREE Full text] [doi: 10.4158/EP15672.GL] [Medline: 25869408]
29. American Diabetes Association. 6. Glycemic Targets. Diabetes Care 2017;40(Supplement 1):S48-S56. [doi: 10.2337/dc17-s009]
30. Qaseem A, Wilt TJ, Kansagara D, Horwitch C, Barry MJ, Forciea MA, Clinical Guidelines Committee of the American College of Physicians. Hemoglobin A1c targets for glycemic control with pharmacologic therapy for nonpregnant adults with type 2 diabetes mellitus: a guidance statement update from The American College of Physicians. Ann Intern Med 2018 Apr 17;168(8):569-576. [doi: 10.7326/M17-0939] [Medline: 29507945]
31. Nazi KM. Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 2010;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]
32. Canadian Agency for Drugs and Technologies in Health. 2014. HbA1c Testing Frequency: A Review of the Clinical Evidence and Guidelines URL: https://www.cadth.ca/hba1c-testing-frequency-review-clinical-evidence-and-guidelines [accessed 2020-03-12]
33. Cho WH, Lee H, Kim C, Lee S, Choi K. The impact of visit frequency on the relationship between service quality and outpatient satisfaction: a South Korean study. Health Serv Res 2004 Feb;39(1):13-33 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00213.x] [Medline: 14965075]
34. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008 Oct 9;359(15):1577-1589. [doi: 10.1056/NEJMoa0806470] [Medline: 18784090]
35. Amiel SA, Sherwin RS, Simonson DC, Tamborlane WV. Effect of intensive insulin therapy on glycemic thresholds for counterregulatory hormone release. Diabetes 1988 Jul;37(7):901-907. [doi: 10.2337/diab.37.7.901] [Medline: 3290007]
36. Wong JC, Dolan LM, Yang TT, Hood KK. Insulin pump use and glycemic control in adolescents with type 1 diabetes: Predictors of change in method of insulin delivery across two years. Pediatr Diabetes 2015 Dec;16(8):592-599 [FREE Full text] [doi: 10.1111/pedi.12221] [Medline: 25387433]

37. Wheeler BJ, Heels K, Donaghue KC, Reith DM, Ambler GR. Insulin pump-associated adverse events in children and adolescents--a prospective study. Diabetes Technol Ther 2014 Sep;16(9):558-562. [doi: 10.1089/dia.2013.0388] [Medline: 24796368]
38. Kordonouri O, Biester T, Schnell K, Hartmann R, Tsioli C, Fath M, et al. Lipoatrophy in children with type 1 diabetes: an increasing incidence? J Diabetes Sci Technol 2015 Mar;9(2):206-208 [FREE Full text] [doi: 10.1177/1932296814558348] [Medline: 25411060]
39. Yeh H, Brown TT, Maruthur N, Ranasinghe P, Berger Z, Suh YD, et al. Comparative effectiveness and safety of methods of insulin delivery and glucose monitoring for diabetes mellitus: a systematic review and meta-analysis. Ann Intern Med 2012 Sep 4;157(5):336-347. [doi: 10.7326/0003-4819-157-5-201209040-00508] [Medline: 22777524]
40. Cho J, Chang S, Kwon H, Choi Y, Ko S, Moon S, et al. Long-term effect of the internet-based glucose monitoring system on HbA1c reduction and glucose stability: a 30-month follow-up study for diabetes management with a ubiquitous medical care system. Diabetes Care 2006 Dec;29(12):2625-2631. [doi: 10.2337/dc05-2371] [Medline: 17130195]
41. Jeon JY, Ko S, Kwon H, Kim NH, Kim JH, Kim CS, Taskforce Team of Diabetes Fact Sheet of the Korean Diabetes Association. Prevalence of Diabetes and Prediabetes according to Fasting Plasma Glucose and HbA1c. Diabetes Metab J 2013 Oct;37(5):349-357 [FREE Full text] [doi: 10.4093/dmj.2013.37.5.349] [Medline: 24199164]
42. Rajalakshmi R, Prathiba V, Mohan V. Does tight control of systemic factors help in the management of diabetic retinopathy? Indian J Ophthalmol 2016 Jan;64(1):62-68 [FREE Full text] [doi: 10.4103/0301-4738.178146] [Medline: 26953026]
43. Conway BN, Long DM, Figaro MK, May ME. Glycemic control and fracture risk in elderly patients with diabetes. Diabetes Res Clin Pract 2016 May;115:47-53 [FREE Full text] [doi: 10.1016/j.diabres.2016.03.009] [Medline: 27242122]
44. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and impact of real-world clinical data for the practicing clinician. Adv Ther 2018 Nov;35(11):1763-1774 [FREE Full text] [doi: 10.1007/s12325-018-0805-y] [Medline: 30357570]
45. Jeon E, Park H. Development of the IMB model and an evidence-based diabetes self-management mobile application. Healthc Inform Res 2018 Apr;24(2):125-138 [FREE Full text] [doi: 10.4258/hir.2018.24.2.125] [Medline: 29770246]

Abbreviations

ADA: American Diabetes Association
AMC: Asan Medical Center
ANCOVA: analysis of covariance
DCSI: diabetes complications severity index
EMR: electronic medical record
HbA$_{1c}$: hemoglobin A$_{1c}$
ICD-10: International Classification of Diseases 10th Revision
IRB: institutional review board
KDA: Korean Diabetes Association
MCMH: My Chart in My Hand
mHealth: mobile health
mPHR: mobile personal health record
PGHD: patient-generated health data
PHR: personal health record
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 05.07.19; peer-reviewed by SY Jung, L Luo, S Ross; comments to author 16.12.19; revised version received 10.02.20; accepted 24.02.20; published 02.06.20.

Please cite as:
Seo D, Park YR, Lee Y, Kim JY, Park JY, Lee JH
The Use of Mobile Personal Health Records for Hemoglobin A1c Regulation in Patients With Diabetes: Retrospective Observational Study
J Med Internet Res 2020;22(6):e15372
URL: https://www.jmir.org/2020/6/e15372
doi:10.2196/15372
PMID:32484447

Original Paper

Comparisons Between Hypothesis- and Data-Driven Approaches for Multimorbidity Frailty Index: A Machine Learning Approach

Li-Ning Peng[1,2,3*], MD, PhD; Fei-Yuan Hsiao[4,5,6*], PhD; Wei-Ju Lee[1,2,3,7], MD, PhD; Shih-Tsung Huang[4], PharmD, MS; Liang-Kung Chen[1,2,3], MD, PhD

[1]Aging and Health Research Center, National Yang Ming University, Taipei, Taiwan
[2]Department of Geriatric Medicine, National Yang Ming University School of Medicine, Taipei, Taiwan
[3]Center for Geriatrics and Gerontology, Taipei Veterans General Hospital, Taipei, Taiwan
[4]Graduate Institute of Clinical Pharmacy, College of Medicine, National Taiwan University, Taipei, Taiwan
[5]School of Pharmacy, College of Medicine, National Taiwan University, Taipei, Taiwan
[6]Department of Pharmacy, National Taiwan University Hospital, Taipei, Taiwan
[7]Department of Family Medicine, Taipei Veterans General Hospital Yuanshan Branch, Yi-Lan, Taiwan
[*]these authors contributed equally

Corresponding Author:
Liang-Kung Chen, MD, PhD
Center for Geriatrics and Gerontology
Taipei Veterans General Hospital
201, Section 2, Shih-Pai Road
Taipei, 11217
Taiwan
Phone: 886 2 28757830
Email: lkchen2@vghtpe.gov.tw

Abstract

Background: Using big data and the theory of cumulative deficits to develop the multimorbidity frailty index (mFI) has become a widely accepted approach in public health and health care services. However, constructing the mFI using the most critical determinants and stratifying different risk groups with dose-response relationships remain major challenges in clinical practice.

Objective: This study aimed to develop the mFI by using machine learning methods that select variables based on the optimal fitness of the model. In addition, we aimed to further establish 4 entities of risk using a machine learning approach that would achieve the best distinction between groups and demonstrate the dose-response relationship.

Methods: In this study, we used Taiwan's National Health Insurance Research Database to develop a machine learning multimorbidity frailty index (ML-mFI) using the theory of cumulative diseases/deficits of an individual older person. Compared to the conventional mFI, in which the selection of diseases/deficits is based on expert opinion, we adopted the random forest method to select the most influential diseases/deficits that predict adverse outcomes for older people. To ensure that the survival curves showed a dose-response relationship with overlap during the follow-up, we developed the distance index and coverage index, which can be used at any time point to classify the ML-mFI of all subjects into the categories of fit, mild frailty, moderate frailty, and severe frailty. Survival analysis was conducted to evaluate the ability of the ML-mFI to predict adverse outcomes, such as unplanned hospitalizations, intensive care unit (ICU) admissions, and mortality.

Results: The final ML-mFI model contained 38 diseases/deficits. Compared with conventional mFI, both indices had similar distribution patterns by age and sex; however, among people aged 65 to 69 years, the mean mFI and ML-mFI were 0.037 (SD 0.048) and 0.0070 (SD 0.0254), respectively. The difference may result from discrepancies in the diseases/deficits selected in the mFI and the ML-mFI. A total of 86,133 subjects aged 65 to 100 years were included in this study and were categorized into 4 groups according to the ML-mFI. Both the Kaplan-Meier survival curves and Cox models showed that the ML-mFI significantly predicted all outcomes of interest, including all-cause mortality, unplanned hospitalizations, and all-cause ICU admissions at 1, 5, and 8 years of follow-up (*P*<.01). In particular, a dose-response relationship was revealed between the 4 ML-mFI groups and adverse outcomes.

Conclusions: The ML-mFI consists of 38 diseases/deficits that can successfully stratify risk groups associated with all-cause mortality, unplanned hospitalizations, and all-cause ICU admissions in older people, which indicates that precise, patient-centered medical care can be a reality in an aging society.

(J Med Internet Res 2020;22(6):e16213) doi:10.2196/16213

KEYWORDS

multimorbidity frailty index; machine learning; random forest; unplanned hospitalizations; intensive care unit admissions; mortality

Introduction

Population aging is a global phenomenon that poses various challenges to societies [1]. The health characteristics of older people and their health care service utilization differ greatly from those of younger adults [2], and frailty plays a pivotal role in the health of older people [3-5]. Frailty has been widely accepted as a geriatric syndrome that substantially increases the complexity of diseases and the burden of care [3-5]. In addition, frailty is recognized as an intermediate state between healthy and unhealthy states, and the potential reversibility of its nature highlights the importance of considering frailty when aiming to maintain the health of older people [6]. Moreover, frailty involves the coexistence of multiple comorbid conditions, such as polypharmacy, depression, cognitive impairment, falls, and malnutrition [7]. Therefore, the early identification of frailty and appropriate intervention remain the core of health care services for older people.

Despite the clinical significance of frailty, conceptual and operational definitions of frailty are inconsistent across studies [8]. Currently, the two most widely accepted approaches include the phenotypic approach for physical frailty and the frailty index based on the theory of cumulative deficits [9]. Although the definitions of frailty provided by the two approaches overlapped to some extent, the major discrepancy is in the prefrail group, such that physically prefrail subjects demonstrated a wide range on the frailty index. Nevertheless, both definitions remain the most widely accepted [10]. The theory of cumulative deficits proposed that aging may be characterized by the presence of cumulative deficits in various domains of health (eg, multimorbidity, functional assessment, and psychosocial perspectives) [9]. With a sufficient number of variables, the individual component of the frailty index was considered the same weight to constitute the frailty index. Researchers applied the theory of cumulative deficits to various data sets and validated the ability of the frailty index (FI) to predict adverse clinical outcomes [4,9]. Internationally, documented health care services data sets have been widely used to develop the FI for the prediction of health outcomes, and studies from different countries have all shown optimal results [5,11,12]. In the United Kingdom, researchers developed the electronic FI (eFI) using electronic medical records, which significantly predicted the mortality of older people [13,14]. Using similar principles, we developed the multimorbidity FI (mFI) using Taiwan’s National Health Insurance data set and significantly predicted mortality, hospitalizations, and admissions to critical care units [4]. However, it is always challenging to use data sets with large study samples and many variables to select appropriate variables to construct an FI and to optimally categorize the FI into risk classes. Both eFI and mFI adopted expert recommendations in the selection of variables, and the eFI and mFI were then categorized into quartiles for group comparisons, which is a widely accepted approach. Nonetheless, selecting variables based on expert recommendations may result in a failure to recognize previously unidentified associations. In addition, the quartile approach for risk group categorization may successfully be used to construct the prediction model, but the intergroup comparisons in survival analysis may overlap and fail to establish a clear distinction.

Therefore, this study aimed to develop the mFI by using machine learning methods that select variables based on the best fitness of the model. Furthermore, we aim to further establish 4 entities of risk using a machine learning approach and ensure the dose-response relationship and the best distinction between groups.

Methods

Study Design and Participants

This is a retrospective cohort study using data from Taiwan's National Health Insurance Research Database (NHIRD). Details about the NHIRD have been published [15]. Briefly, the NHIRD is a nationwide database composed of outpatient and inpatient claims, and it covers more than 99% of Taiwan's population. The data are checked for quality and maintained by the Data Science Centre of the Ministry of Health and Welfare of Taiwan. We used a subset of the NHIRD, which contains claims data for one million randomly selected beneficiaries from the Registry of Beneficiaries of the NHIRD in 2005. The study cohort consisted of 86,133 older adults aged 65 to 100 years who had full National Health Insurance (NHI) coverage from January 1, 2005, to December 31, 2005. Claims data from 2005 to 2013 for the one million beneficiaries was extracted to compose a 9-year (2005-2013) panel of claims for analysis. The study protocol was approved by the Research Ethics Committee of the National Taiwan University Hospital (NTUH-REC-201403069W).

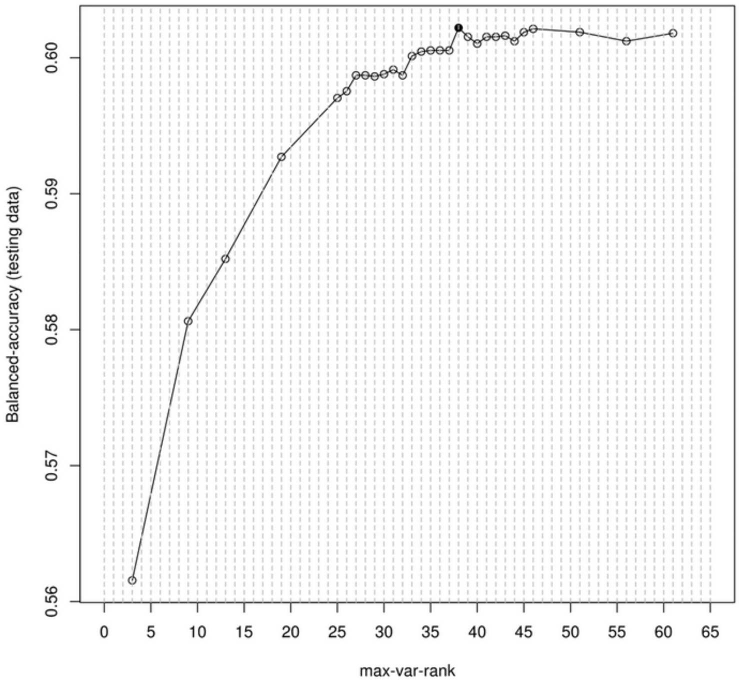
Construction of the Machine Learning–Based Multimorbidity Frailty Index

The mFI was constructed following standard procedures [16], and this method has been validated in the Taiwanese population [4,5]. Disease diagnoses (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM]) from outpatient and inpatient claims of the NHIRD between January 1 and December 31, 2005, were used to identify accumulated deficits to construct mFI. We adopted an algorithm widely used in studies using NHIRD as the data source to validate the diagnostic codes of the specified deficits within

NHIRD; that is, only those who had at least 3 outpatient claim records or 1 inpatient claim record for that specified diagnosis code were considered to have the specified deficit. For example, an older adult must have at least 3 outpatient claim records or 1 inpatient claim record of diabetes mellitus [ICD-9-CM: 250] to be defined as having a deficit based on our definition.

A random forest method, with significant improvements in classification accuracy that resulted from growing an ensemble of trees and letting them vote for the most popular class, was adopted [17]. The variable importance of the random forest uses mean decrease accuracy to determine the specific conditions of machine learning–based multimorbidity frailty index (ML-mFI). The adequate constructive number of ML-mFI was 38 conditions, when the model accuracy reached the highest level, 0.602 (Figure 1 and Multimedia Appendix 1). The ML-mFI was calculated as the number of conditions a person encountered in a year out of the 38 selected ones.

Figure 1. Numbers of diseases versus random forest model accuracy to determine an adequate number of frailty indexes.



Determination of Frailty Status by ML-mFI

All subjects were further categorized into 4 entities (fit, mild frailty, moderate frailty, and severe frailty) based on their risk status; this categorization was used by a previous study [4]. The fundamental rules for risk stratification included the following: (1) the individual risk groups were significantly different from each other, and (2) the health risk of these groups showed a dose-response relationship (ie, those in the severe frailty group had a higher risk than those in the moderate frailty group, who had a higher risk than those in the mild frailty group, and so on, at any follow-up time point after the first year). To achieve this purpose, we developed two indices, the distance index and the coverage index, which ensured the distinction and dose-response relationship of all survival curves.

The distance index measured the distance between each survival curve and the stability of those distances within groups. At any time point, the distance index was defined as . Therefore, the distances within groups are wider and more stable when the distance index is larger (Multimedia Appendix 2). Conversely, the coverage index aimed to evaluate the length of the confidence interval for each survival curve. The total length of the confidence intervals indicated the overall estimated error in the grouping method. In Multimedia Appendix 3, the coverage index was defined as at any individual time point, where L_total measured the difference in the estimated survival probability between the fit group and the severe frailty group, and L_error measured the total estimated errors within the 4 groups. When the coverage index is smaller, the estimated error within groups is smaller. With the application of both the distance index and the coverage index, the levels of frailty were successfully categorized into 4 groups by values of ML-mFI: fit was indicated by 0≤ML-mFI<0.026; mild frailty was 0.026≤ML-mFI<0.105; moderate frailty was 0.105≤ML-mFI<0.157; and severe frailty was 0.157≤ML-mFI. In the survival analysis, the grouping strategy successfully categorized all subjects into 4 groups with significant distinction during the follow-up period. In other words, there were no overlaps between the survival curves and the dose-response relationship between groups was clearly shown.

Outcomes of Interest

The outcomes of interest in this study include all-cause mortality, unplanned hospitalizations, and intensive care unit (ICU) admissions. The date of mortality was identified as the date of disenrollment from the NHIRD, which has been validated in a previous study [4]. Unplanned hospitalizations were any unexpected hospitalizations after an emergency department visit. ICU admissions were any hospital admissions with the use of ICU services. All study subjects were continuously followed from January 1, 2006, to the occurrence of each outcome or the end of 2013, whichever came first. For the outcomes of unplanned hospitalizations and ICU admissions, subjects were censored at death if it occurred first. Preplanned analyses were conducted to evaluate the effectiveness of ML-mFI in predicting outcomes at 1, 5, and 8 years.

Statistical Analysis

Numerical variables were expressed as the mean (SD), and categorical variables were expressed as a number or percentage. A random forest method not only determined the number of disease items comprising ML-mFI but also identified potential conditions of ML-mFI with prediction accuracy and variable importance. The distance index and coverage index with min-max and max-min criteria were used to determine cut points and categorize the frailty group by ML-mFI automatically. The Kaplan-Meier survival curve with the log-rank test was used to examine the association between categories of ML-mFI (fit, mild frailty, moderate frailty, and severe frailty) and 8-year mortality and hospitalizations. Cox proportional hazard models were used to estimate the hazard ratios (HRs) and 95% CIs for mortality and hospitalizations at 1, 5, and 8 years after the ML-mFI and mFI were estimated (based on a previous study [4]), considering both to be the independent variable. We further included age and gender as covariates in all adjusted models. Sex-specific analysis was conducted.

All of the analyses were performed using R Version 3.4.4 (R Foundation for Statistical Computing). A two-sided *P* value of <.05 was considered statistically significant. The coxph function in the survival package showed nonviolation of the proportional hazards assumption and a linear relationship between the log hazard and each covariate. The random forest and importance functions in the randomForest package showed the model building and variable importance to predict the outcome occurrence and comprise ML-mFI, respectively.

Results

Construction of ML-mFI

The final ML-mFI with the highest model accuracy (0.6022061) contained 38 conditions (Multimedia Appendix 1). Details of convergences and divergences of composing conditions among ML-mFI and mFI are shown in Multimedia Appendix 4. Table 1 compares the ML-mFI group and traditional mFI by age and sex. There were two similar distribution patterns on mFI and ML-mFI. ML-mFI increased with age, but reached a plateau at age 80 years and older. Both indices were higher in males, which is compatible with the shorter life expectancy of men in Taiwan. However, the mFI was calculated based on 32 selected conditions a person may have in a year, while the ML-mFI was calculated based on 38 selected conditions a person may have in a year; thus, the actual numbers on the mFI and ML-mFI were very different. Among people aged 65 to 69 years, the mean mFI and ML-mFI were 0.037 (SD 0.048) and 0.0070 (SD 0.0254), respectively. The difference may result from discrepancies in the conditions selected on the mFI and the ML-mFI. For example, some conditions were selected only on the ML-mFI but not on the mFI (eg, ICD-9-CM: 250 [diabetes mellitus] and, vice versa, ICD-9-CM: 374 [entropion]). These discrepancies have been shown in Multimedia Appendix 1.

Table 1. Comparisons of mFI and ML-mFI by age and sex.[a,b]

Age (years)	All subjects (N=86,133)		Male (n=42,914)		Female (n=43,219)	
	mFI, mean (SD)	ML-mFI, mean (SD)	mFI, mean (SD)	ML-mFI, mean (SD)	mFI, mean (SD)	ML-mFI, mean (SD)
65-69 (n=28,480)	0.037 (0.048)	0.0070 (0.0254)	0.038 (0.049)	0.0076 (0.0264)	0.037 (0.046)	0.0065 (0.0246)
70-74 (n=23,700)	0.050 (0.056)	0.0106 (0.0322)	0.053 (0.060)	0.0115 (0.0339)	0.046 (0.053)	0.0096 (0.0304)
75-79 (n=18,765)	0.062 (0.065)	0.0150 (0.0400)	0.067 (0.070)	0.0160 (0.0417)	0.056 (0.059)	0.0138 (0.0379)
80-84 (n=9934)	0.070 (0.071)	0.0201 (0.0473)	0.076 (0.075)	0.0212 (0.0490)	0.064 (0.065)	0.0190 (0.0455)
≥85 (n=5254)	0.070 (0.074)	0.0234 (0.0505)	0.077 (0.080)	0.0245 (0.0531)	0.064 (0.069)	0.0224 (0.0483)
Total (N=86,133)	0.052 (0.060)	0.0122 (0.0359)	0.056 (0.064)	0.0132 (0.0376)	0.048 (0.056)	0.0113 (0.0341)

[a] mFI: multimorbidity frailty index.

[b] ML-mFI: machine learning multimorbidity frailty index.

Survival Analysis

Overall, 86,133 subjects aged 65 to 100 years were included in this study. With a mean follow-up of 6.57 (SD 2.37) years, 30,136 deaths (34.99%) occurred among the study cohort during the study period. Figure 2 summarizes the results of the Kaplan-Meier survival curves estimating 4 levels of ML-mFI on all-cause mortality, unplanned hospitalization, and ICU admission, and shows that ML-mFI significantly predicted all these outcomes of interest.

Table 2 shows the hazard ratios of all-cause mortality, unplanned admissions, and ICU admissions for the ML-mFI and the mFI at the 1-, 5- and 8-year follow-up periods. Among all three outcomes of interest, ML-mFI posed higher hazards than did mFI. For example, those who were categorized as severely frail by the mFI or the ML-mFI were associated with 4.97-fold (adjusted HR 4.97, 95% CI 4.49-5.50) and 11.4-fold (adjusted HR 11.40, 95% CI 10.32-12.59) increases in 1-year all-cause mortality, respectively. Similar patterns were observed for 5-year and 8-year all-cause mortality.

Figure 2. The 8-year Kaplan-Meier survival curve for the outcome of (A) all-cause mortality, (B) unplanned hospitalizations, and (C) intensive care unit admissions for different frailty categories.

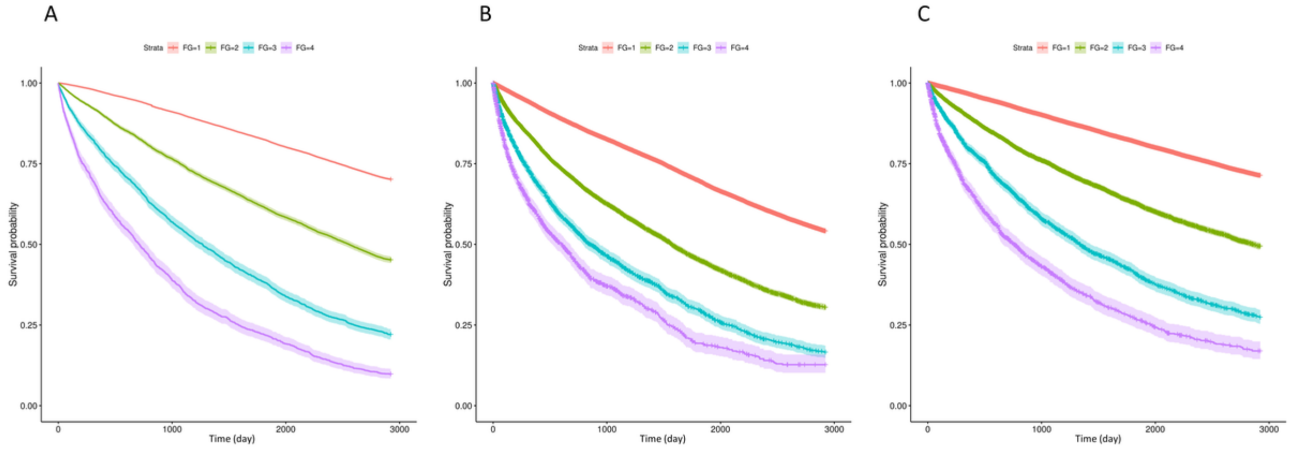


Table 2. Hazard ratios of all-cause mortality, unplanned hospitalizations, and intensive care unit admissions for the ML-mFI and the mFI at the 1-, 5- and 8-year follow-up periods.[a,b,c] All values are given as hazard ratio (95% CI).

Adverse outcomes at follow-up periods	Mild frailty		Moderate frailty		Severe frailty	
	mFI (n=14,244)	ML-mFI (n=9366)	mFI (n=4741)	ML-mFI (n=2522)	mFI (n=2498)	ML-mFI (n=1488)
1-year all-cause mortality HR[d]						
Unadjusted	2.21 (2.04-2.39)	3.66 (3.38-3.97)	4.09 (3.72-4.50)	8.81 (8.00-9.71)	7.52 (6.81-8.30)	16.62 (15.08-18.32)
Adjusted	1.86 (1.71-2.01)	3.13 (2.89-3.39)	3.08 (2.80-3.39)	6.79 (6.15-7.49)	4.97 (4.49-5.50)	11.40 (10.32-12.59)
5-year all-cause mortality HR						
Unadjusted	1.76 (1.70-1.82)	2.57 (2.48-2.67)	2.85 (2.72-2.99)	5.27 (5.00-5.55)	5.00 (4.74-5.28)	9.02 (8.49-9.58)
Adjusted	1.46 (1.41-1.52)	2.19 (2.11-2.27)	2.14 (2.04-2.25)	4.04 (3.83-4.26)	3.28 (3.11-3.46)	6.15 (5.79-6.54)
8-year all-cause mortality HR						
Unadjusted	1.69 (1.64-1.74)	2.32 (2.25-2.39)	2.65 (2.55-2.76)	4.72 (4.54-4.94)	4.50 (4.29-4.71)	8.05 (7.61-8.51)
Adjusted	1.41 (1.37-1.45)	1.99 (1.93-2.05)	2.01 (1.93-2.09)	3.70 (3.53-3.88)	2.98 (2.84-3.12)	5.52 (5.22-5.84)
1-year unplanned hospitalization HR						
Unadjusted	2.08 (1.97-2.20)	2.86 (2.70-3.02)	3.30 (3.07-3.54)	5.21 (4.82-5.64)	5.29 (4.88-5.73)	7.65 (6.99-8.38)
Adjusted	1.91 (1.80-2.01)	2.63 (2.49-2.79)	2.85 (2.65-3.06)	4.53 (4.18-4.90)	4.28 (3.94-4.64)	6.20 (5.66-6.80)
5-year unplanned hospitalization HR						
Unadjusted	1.78 (1.73-1.83)	2.28 (2.21-2.36)	2.51 (2.40-2.62)	3.79 (3.59-4.00)	3.85 (3.65-4.06)	5.43 (5.07-5.83)
Adjusted	1.61 (1.57-1.66)	2.09 (2.02-2.16)	2.14 (2.05-2.24)	3.23 (3.06-3.41)	3.05 (2.89-3.23)	4.33 (4.04-4.65)
8-year unplanned hospitalization HR						
Unadjusted	1.67 (1.63-1.71)	2.11 (2.05-2.17)	2.32 (2.24-2.41)	3.53 (3.36-3.71)	3.53 (3.36-3.71)	5.03 (4.69-5.38)
Adjusted	1.51 (1.48-1.55)	1.93 (1.87-1.99)	1.98 (1.91-2.06)	3.01 (2.86-3.17)	2.79 (2.65-2.94)	3.98 (3.72-4.27)
1-year intensive care unit admission HR						
Unadjusted	2.34 (2.18-2.52)	3.23 (3.00-3.48)	4.32 (3.95-4.72)	6.70 (6.08-7.38)	7.04 (6.38-7.76)	12.16 (10.98-13.46)
Adjusted	2.09 (1.94-2.25)	2.91 (2.70-3.13)	3.59 (3.28-3.92)	5.64 (5.11-6.23)	5.35 (4.84-5.91)	9.41 (8.49-10.44)
5-year intensive care unit admission HR						
Unadjusted	1.86 (1.79-1.93)	2.39 (2.30-2.48)	2.92 (2.78-3.07)	4.75 (4.48-5.04)	4.84 (4.56-5.14)	7.84 (7.30-8.42)
Adjusted	1.64 (1.58-1.70)	2.14 (2.06-2.23)	2.42 (2.30-2.54)	3.96 (3.74-4.21)	3.65 (3.43-3.87)	6.00 (5.58-6.44)
8-year intensive care unit admission HR						
Unadjusted	1.74 (1.69-1.79)	2.20 (2.12-2.28)	2.69 (2.58-2.81)	4.35 (4.12-4.59)	4.28 (4.05-4.52)	7.10 (6.63-7.60)
Adjusted	1.54 (1.49-1.59)	1.98 (1.92-2.05)	2.23 (2.14-2.34)	3.68 (3.49-3.89)	3.24 (3.06-3.42)	5.46 (5.10-5.85)

[a] For all outcomes, the comparator is the study subjects in the fit categories (n=64,650). All data were adjusted for age and gender.

[b] ML-mFI: machine learning multimorbidity frailty index.

[c] mFI: multimorbidity frailty index.

[d] HR: hazard ratio.

For unplanned hospitalizations, those who were categorized as severely frail by the mFI or the ML-mFI were associated with 4.28-fold (adjusted HR 4.28, 95% CI 3.94-4.64) and 6.20-fold (adjusted HR 6.20, 95% CI 5.66-6.80) increases in 1-year unplanned hospitalizations, respectively. Similar patterns were observed for 5-year and 8-year all-cause unplanned hospitalizations.

For ICU admissions, those who were categorized as severely frail by the mFI or the ML-mFI were associated with 4.28-fold (adjusted HR 5.35, 95% CI 4.84-5.91) and 9.41-fold (adjusted HR 9.41, 95% CI 8.49-10.44) increases in 1-year ICU admissions, respectively. Similar patterns were observed for 5-year and 8-year all-cause ICU admissions.

Sex-specific analysis showed that both indices were higher in men than in women for various outcomes and follow-up periods (Tables 3 and 4 for males and females, respectively). For example, men in the severe frailty group (as defined by the ML-mFI) were associated with a 12.64-fold increased risk of 1-year mortality, while women in the severe frailty group (as defined by the mFI) were associated with a 10.37-fold increased risk of 1-year mortality.

Table 3. Male hazard ratios of all-cause mortality, unplanned hospitalization, and intensive care unit admission for ML-mFI and mFI among 1-, 5-, and 8-year follow-up periods.[a,b,c]

Adverse outcome and follow-up period	Mild frailty		Moderate frailty		Severe frailty	
	mFI (n=14,244)	ML-mFI (n=9366)	mFI (n=4,741)	ML-mFI (n=2522)	mFI (n=2498)	ML-mFI (n=1488)
All-cause mortality						
1-year	1.83 (1.65-2.04)	3.71 (3.37-4.09)	2.70 (2.37-3.07)	7.67 (6.76-8.69)	4.84 (4.26-5.49)	12.64 (11.20-14.27)
5-year	1.41 (1.35-1.48)	2.53 (2.42-2.65)	1.93 (1.82-2.06)	4.60 (4.29-4.93)	3.07 (2.86-3.28)	6.92 (6.40-7.48)
8-year	1.35 (1.30-1.41)	2.29 (2.21-2.38)	1.85 (1.75-1.95)	4.19 (3.94-4.47)	2.77 (2.61-2.94)	6.27 (5.83-6.74)
Unplanned hospitalization						
1-year	1.87 (1.73-2.01)	2.83 (2.64-3.04)	2.73 (2.48-3.00)	4.90 (4.41-5.45)	4.24 (3.83-4.71)	6.34 (5.63-7.14)
5-year	1.58 (1.51-1.64)	2.25 (2.16-2.35)	2.05 (1.93-2.17)	3.49 (3.25-3.76)	3.00 (2.80-3.21)	4.59 (4.19-5.03)
8-year	1.48 (1.43-1.53)	2.07 (1.99-2.15)	1.91 (1.81-2.01)	3.30 (3.08-3.54)	2.76 (2.59-2.95)	4.29 (3.92-4.69)
Intensive care unit admission						
1-year	2.02 (1.83-2.23)	3.35 (3.06-3.67)	3.28 (2.91-3.69)	5.91 (5.18-6.74)	4.85 (4.27-5.51)	9.50 (8.31-10.86)
5-year	1.58 (1.50-1.66)	2.42 (2.31-2.55)	2.24 (2.10-2.40)	4.33 (4.00-4.69)	3.39 (3.15-3.66)	6.28 (5.72-6.89)
8-year	1.48 (1.42-1.54)	2.22 (2.13-2.32)	2.05 (1.94-2.18)	4.04 (3.75-4.35)	2.99 (2.79-3.21)	5.80 (5.30-6.35)

[a] For all outcomes, the comparator is subjects in fit categories (n=64,650). All data were adjusted for age and gender.

[b] ML-mFI: machine learning multimorbidity frailty index.

[c] mFI: multimorbidity frailty index.

Table 4. Female hazard ratios of all-cause mortality, unplanned hospitalization, and intensive care unit admission for ML-mFI and mFI among 1-, 5-, and 8-year follow-up periods.[a,b,c]

Adverse outcome and follow-up period	Mild frailty		Moderate frailty		Severe frailty	
	mFI (n=14,244)	ML-mFI (n=9366)	mFI (n=4741)	ML-mFI (n=2522)	mFI (n=2498)	ML-mFI (n=1488)
All-cause mortality						
1-year	1.88 (1.66-2.13)	2.56 (2.28-2.87)	3.73 (3.22-4.32)	5.95 (5.19-6.82)	5.29 (4.46-6.27)	10.37 (8.97-12.00)
5-year	1.54 (1.46-1.62)	1.87 (1.77-1.97)	2.52 (2.34-2.71)	3.51 (3.25-3.78)	3.79 (3.46-4.15)	5.52 (5.04-6.05)
8-year	1.48 (1.42-1.55)	1.72 (1.65-1.80)	2.31 (2.17-2.46)	3.24 (3.04-3.46)	3.48 (3.22-3.76)	4.89 (4.49-5.32)
Unplanned hospitalization						
1-year	1.95 (1.80-2.11)	2.45 (2.26-2.65)	3.03 (2.72-3.38)	4.17 (3.72-4.66)	4.36 (3.81-4.98)	6.19 (5.40-7.09)
5-year	1.66 (1.59-1.73)	1.94 (1.86-2.03)	2.28 (2.14-2.44)	2.98 (2.77-3.21)	3.16 (2.89-3.46)	4.12 (3.71-4.58)
8-year	1.56 (1.50-1.62)	1.81 (1.74-1.88)	2.10 (1.98-2.22)	2.74 (2.55-2.94)	2.85 (2.62-3.11)	3.71 (3.34-4.12)
Intensive care unit admission						
1-year	2.18 (1.95-2.44)	2.48 (2.22-2.76)	4.09 (3.56-4.70)	5.39 (4.71-6.16)	6.44 (5.49-7.55)	9.73 (8.40-11.27)
5-year	1.73 (1.64-1.82)	1.88 (1.77-1.99)	2.70 (2.50-2.91)	3.62 (3.33-3.93)	4.21 (3.82-4.65)	5.89 (5.29-6.56)
8-year	1.62 (1.55-1.70)	1.77 (1.69-1.86)	2.54 (2.37-2.71)	3.34 (3.10-3.61)	3.79 (3.47-4.15)	5.27 (4.76-5.85)

[a] For all outcomes, the comparator is subjects in fit categories (n=64,650). All data were adjusted for age and gender.

[b] ML-mFI: machine learning multimorbidity frailty index.

[c] mFI: multimorbidity frailty index.

Discussion

In this study, we successfully used a machine learning approach to define ML-mFI. Specifically, we selected disease/deficit items by the random forest method and ranked the importance of each individual disease accordingly. The selection of these diseases/deficits items to construct ML-mFI was driven solely by data, while the conventional mFI included disease/deficit items based on expert recommendations. Moreover, the combined use of the distance index and coverage index successfully distinguished 4 groups with dose-response risks of adverse outcomes. In epidemiological studies, researchers have often encountered similar challenges in selecting appropriate variables for analysis and optimally categorizing continuous variables into categorical variables for further comparisons. Traditionally, researchers need to search for literature support or adopt a generic approach to develop an optimal statistical model for data interpretation [18-20]. The hypothesis-driven approach for a research question is of great importance in scientific development; however, previously unknown or unidentified factors may be overlooked in the analysis, which may lower the statistical power in the interpretation of the phenomenon. Compared to our previous work where we used a hypothesis-driven approach to construct the mFI [4], the machine learning model selected significantly different disease/deficit items for ML-mFI construction. The traditional approach selected the diseases/deficits of older adults based on the selection criteria, and the machine learning approach identified more disease/deficit items, including chronic diseases, infectious diseases, and even some cancers, but these items did not comprise the majority of disease/deficit items.

The FI developed by Rockwood et al [18] hypothesized that cumulative deficits in various health domains may represent the process of biological aging, and this FI has been widely validated to predict adverse health events and mortality in different countries [5,11,12]. In theory, an FI may consist of as many variables as possible, so there are no issues regarding variable selection. However, to meet the needs of the busy clinical environment, the mFI is derived from the concept of the FI; a selection of age-related chronic conditions were the key variables used to construct the prediction model. Existing studies have shown that these previously developed mFIs can significantly predict the mortality of older adults [13.14]. However, to maximize the effectiveness of the prediction model, using a data-driven approach to construct the ML-mFI may provide better prediction accuracy. Moreover, in the survival analysis, the dose-response relationship is usually expected when grouping continuous measurements into distinct risk groups in association with outcomes. However, the distinction between individual groups from the continuous measurements is not always statistically significant even though the whole model reached statistical significance. For example, in one Taiwanese study, the developed FI was found to predict the adverse outcomes of older adults, which was in line with most related studies [5]. However, in that study, different risk groups that were categorized based on FI tertiles resulted in overlapping survival curves of the intermediate- and high-risk groups; it failed to achieve the stratification of risk groups. The combined use of the distance index and coverage index developed in this study engenders the ability to address the overlapping phenomenon of survival curves.

Although the mFI we developed adequately predicted adverse outcomes for older adults, the ML-mFI showed relatively higher hazard ratios than did the mFI for all health outcomes. Overall, the data-driven ML-mFI may identify different at-risk populations than the hypothesis-driven mFI. The data-driven approach may disclose the phenomenon of the whole data set [21-23], but the hypothesis-driven approach may provide a better explanation for the observations [24,25]. The data-driven approach may not be superior to the hypothesis-driven approach, since the study purpose and research questions may vary greatly. Although a data-driven approach may usually establish a prediction model with better accuracy, it is difficult to implement intervention programs for the observed phenomenon. Applying the theory of cumulative deficits, a large number of variables may be used to construct the prediction model, but it becomes challenging to further utilize the prediction model with a large number of variables. Therefore, researchers have attempted to reduce the number of variables while maintaining optimal prediction accuracy. Our previous study used factor analysis to reduce the 125 selected variables into 35 factors to improve the clinical application [5]. However, the machine learning approach in this study may play a similar role in reducing the selected number of variables and optimizing prediction accuracy. The main strength of this study was to demonstrate the methodological advance of processing a large data set to select appropriate variables to construct a prediction model and to ensure the distinction of different risk groups with dose-response relationships. This methodological advance may facilitate public health or social sciences research, or interdisciplinary research that uses a large data set with a wide array of data characteristics. In particular, the distance index and coverage index would be of great importance for future research to categorize the results of continuous variables into distinct entities with different health risks. Avoiding the overlap of the survival curves of different risk groups by using the distance index and coverage index is important to strengthen the observed phenomenon and the risk group classification.

Therefore, this ML-mFI demonstrated an automatic approach to predict adverse outcomes in older people, and it can be applied to different populations in different countries. Using the same approach, different diseases can be selected to construct the new ML-mFI in another population to predict adverse outcomes in the corresponding population. For example, we further stratified our study population into 3 subcohorts, including those aged 65 to 75 years, 76 to 85 years, and 85 years and older, and we constructed three kinds of ML-mFI for each age group according to the same automatic machine learning approach and model selection criteria. We found that the total deficit number and composing deficits on the ML-mFI, as well the cut-off points of different frailty statuses, are quite different in distinct age groups. For example, the total deficit numbers on the ML-mFI were 59, 47, and 39 for those who were aged 65 to 75 years, 76 to 85 years, and 85 years and older, respectively. In addition, the composing deficits were different, as displayed in Multimedia Appendix 5. Although the composing deficits of the ML-mFI are different in distinct age groups, all of these ML-mFI can successfully predict all-cause mortality (C index>0.6). These findings are inspiring because they indicate that the same machine learning approach can be used to construct one’s own ML-mFI to fulfill this purpose. Individual diseases may have different clinical impacts in different countries due to diagnosis, treatment, and quality of care. Therefore, the results of this study can be applied to different countries and populations using the same approach to construct their own ML-mFI to meet their needs.

Therefore, our ML-mFI could have clinical implications in public health or in health care administration. For example, in large long-term care facilities management, the administration needs to optimize the admission waiting list through the estimation of the mortality of all residents. On the other hand, in public health settings, the government is able to accurately estimate the health risk of residents in a certain geographic area and to provide optimal health care or palliative care services. Traditionally, these decisions were made based on existing medical knowledge, but a data-driven approach may better predict outcomes and optimize the government's public health policy. In clinical practice, the ML-mFI may enable physicians and families to quantify health risks for optimal care planning. Hence, using available electronic medical records, the ML-mFI can be automatically generated and integrated as part of the medical record to facilitate certain forms of decision-making in care planning.

Despite all the effort that went into this study, there are still some limitations. First, like all data-driven studies, the results of this study could not provide or validate a well-established hypothetical framework due to the nature of machine learning. Second, it remained difficult to develop further intervention programs based on the diagnostic entities identified by machine learning. Third, another data set is needed to examine whether overfitting exists in the machine learning model. Finally, as in most of the previous frailty index studies, although we adjusted for age and sex as covariates in the Cox model, we were unable to access some residual confounders not routinely captured in a claims database, such as disease severity or lifestyle factors (eg, physical activity and diet).

In conclusion, the ML-mFI significantly predicted adverse health outcomes for older adults, and the risk groups defined by the combination of the distance index and coverage index distinguished the different risk groups with dose-response relationships and clear distinctions. The methodological advance of this study also had further research implications for studies with similar data and research questions. The data-driven approach may provide better prediction accuracy than the hypothesis-driven approach, but the superiority of the data-driven approach requires further study for confirmation.

Acknowledgments

The authors thank the National Health Insurance Association and National Health Research Institutes for making available the databases used in this study; the Ministry of Science and Technology, Taiwan, for the funding of this project (MOST 107-2634-F-010-001, MOST 108-2634-F-010-001, MOST 109-2634-F-010-001); and Tsung-Hsien Tsai, Pei-Jung Chen, and ACER for supporting data analysis. However, this paper in no way represents any official position of the National Health Insurance Association, the National Health Research Institutes, the Ministry of Science and Technology, and ACER. The authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

STH, HFY, and CLK received research project funding (MOST 107-2634-F-010-001) sponsored by the Ministry of Science and Technology (Taiwan). All authors have declared no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Multimedia Appendix 1
Model accuracy with various numbers of composite diseases.
[DOCX File , 14 KB - jmir_v22i6e16213_app1.docx]

Multimedia Appendix 2
Survival curves with 4 frail groups to depict the calculation of the distance index.
[DOCX File , 125 KB - jmir_v22i6e16213_app2.docx]

Multimedia Appendix 3
Survival curves with 4 frail groups to depict the calculation of the coverage index.
[DOCX File , 143 KB - jmir_v22i6e16213_app3.docx]

Multimedia Appendix 4
Convergences and divergences between mFI and ML-mFI.
[DOCX File , 15 KB - jmir_v22i6e16213_app4.docx]

Multimedia Appendix 5
Total number of deficits and the composing deficits of the ML-mFI in three subcohorts, including 65 to 75 years old, 76 to 85 years old, and >85 years old.
[DOCX File , 20 KB - jmir_v22i6e16213_app5.docx]

References

1. Chen LK, Inoue H, Won CW, Lin CH, Lin KF, Tsay SF, et al. Challenges of urban aging in Taiwan: Summary of urban aging forum. Journal of Clinical Gerontology and Geriatrics 2013 Dec;4(4):97-101. [doi: 10.1016/j.jcgg.2013.05.001]
2. Lu WH, Lee WJ, Chen LK, Hsiao FY. Comparisons of annual health care utilization, drug consumption, and medical expenditure between the elderly and general population in Taiwan. Journal of Clinical Gerontology and Geriatrics 2016 Jun;7(2):44-47. [doi: 10.1016/j.jcgg.2015.08.002]
3. Lee WJ, Peng LN, Lin CH, Lin HP, Loh CH, Chen LK. The synergic effects of frailty on disability associated with urbanization, multimorbidity, and mental health: implications for public health and medical care. Sci Rep 2018 Sep 20;8(1):14125 [FREE Full text] [doi: 10.1038/s41598-018-32537-5] [Medline: 30237508]
4. Wen YC, Chen LK, Hsiao FY. Predicting mortality and hospitalization of older adults by the multimorbidity frailty index. PLoS One 2017 Nov 16;12(11):e0187825 [FREE Full text] [doi: 10.1371/journal.pone.0187825] [Medline: 29145407]
5. Lin SY, Lee WJ, Chou MY, Peng LN, Chiou SY, Chen LK. Frailty Index Predicts All-Cause Mortality for Middle-Aged and Older Taiwanese: Implications for Active-Aging Programs. PLoS One 2016 Aug 18;11(8):e0161456 [FREE Full text] [doi: 10.1371/journal.pone.0161456] [Medline: 27537684]
6. Van der Elst M, Schoenmakers B, Duppen D, Lambotte D, Fret B, Vaes B, D-SCOPE Consortium. Interventions for frail community-dwelling older adults have no significant effect on adverse outcomes: a systematic review and meta-analysis. BMC Geriatr 2018 Oct 20;18(1):249 [FREE Full text] [doi: 10.1186/s12877-018-0936-7] [Medline: 30342479]
7. Chen LK, Hwang AC, Liu LK, Lee WJ, Peng LN. Frailty Is a Geriatric Syndrome Characterized by Multiple Impairments: A Comprehensive Approach Is Needed. J Frailty Aging 2016;5(4):208-213. [doi: 10.14283/jfa.2016.109] [Medline: 27883167]
8. Lee L, Patel T, Hillier LM, Maulkhan N, Slonim K, Costa A. Identifying frailty in primary care: A systematic review. Geriatr Gerontol Int 2017 Oct 12;17(10):1358-1377. [doi: 10.1111/ggi.12955] [Medline: 28401737]
9. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. J Gerontol A Biol Sci Med Sci 2007 Jul 01;62(7):722-727. [doi: 10.1093/gerona/62.7.722] [Medline: 17634318]
10. Blodgett J, Theou O, Kirkland S, Andreou P, Rockwood K. Frailty in NHANES: Comparing the frailty index and phenotype. Arch Gerontol Geriatr 2015 May;60(3):464-470. [doi: 10.1016/j.archger.2015.01.016] [Medline: 25697060]
11. Hoogendijk EO, Rockwood K, Theou O, Armstrong JJ, Onwuteaka-Philipsen BD, Deeg DJH, et al. Tracking changes in frailty throughout later life: results from a 17-year longitudinal study in the Netherlands. Age Ageing 2018 Sep 01;47(5):727-733. [doi: 10.1093/ageing/afy081] [Medline: 29788032]
12. Tang Z, Wang C, Song X, Shi J, Mitnitski A, Fang X, et al. Co-occurrence of cardiometabolic diseases and frailty in older Chinese adults in the Beijing Longitudinal Study of Ageing. Age Ageing 2013 May;42(3):346-351. [doi: 10.1093/ageing/aft004] [Medline: 23462693]
13. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing 2016 May 03;45(3):353-360 [FREE Full text] [doi: 10.1093/ageing/afw039] [Medline: 26944937]
14. Boyd PJ, Nevard M, Ford JA, Khondoker M, Cross JL, Fox C. The electronic frailty index as an indicator of community healthcare service utilisation in the older population. Age Ageing 2019 Mar 01;48(2):273-277. [doi: 10.1093/ageing/afy181] [Medline: 30590413]
15. Hsiao FY, Yang CL, Huang YT, Huang WF. Using Taiwan's national health insurance research databases for pharmacoepidemiology research. J Food Drug Anal Jul 2007 Jun 15;15(2):99-108.
16. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr 2008 Sep 30;8(1):24 [FREE Full text] [doi: 10.1186/1471-2318-8-24] [Medline: 18826625]
17. Breiman L. Random forests. Machine learning 2001;45(1):5-32. [doi: 10.1201/9780367816377-11]
18. Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ 2005 Aug 30;173(5):489-495 [FREE Full text] [doi: 10.1503/cmaj.050051] [Medline: 16129869]
19. Rockwood K, Blodgett JM, Theou O, Sun MH, Feridooni HA, Mitnitski A, et al. A Frailty Index Based On Deficit Accumulation Quantifies Mortality Risk in Humans and in Mice. Sci Rep 2017 Feb 21;7(1):43068 [FREE Full text] [doi: 10.1038/srep43068] [Medline: 28220898]
20. Rockwood K, Mitnitski A. Frailty defined by deficit accumulation and geriatric medicine defined by frailty. Clin Geriatr Med 2011 Feb;27(1):17-26. [doi: 10.1016/j.cger.2010.08.008] [Medline: 21093719]
21. Shin EK, Mahajan R, Akbilgic O, Shaban-Nejad A. Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits. NPJ Digit Med 2018 Oct 2;1(1):50 [FREE Full text] [doi: 10.1038/s41746-018-0056-y] [Medline: 31304329]
22. Mortazavi BJ, Bucholz EM, Desai NR, Huang C, Curtis JP, Masoudi FA, et al. Comparison of Machine Learning Methods With National Cardiovascular Data Registry Models for Prediction of Risk of Bleeding After Percutaneous Coronary Intervention. JAMA Netw Open 2019 Jul 03;2(7):e196835 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.6835] [Medline: 31290991]

23. Zador Z, Landry A, Cusimano MD, Geifman N. Multimorbidity states associated with higher mortality rates in organ dysfunction and sepsis: a data-driven analysis in critical care. Crit Care 2019 Jul 08;23(1):247 [FREE Full text] [doi: 10.1186/s13054-019-2486-6] [Medline: 31287020]
24. Amaral R, Pereira AM, Jacinto T, Malinovschi A, Janson C, Alving K, et al. Comparison of hypothesis- and data-driven asthma phenotypes in NHANES 2007-2012: the importance of comprehensive data availability. Clin Transl Allergy 2019 Mar 13;9(1):17 [FREE Full text] [doi: 10.1186/s13601-019-0258-7] [Medline: 30918624]
25. Previdelli AN, de Andrade SC, Fisberg RM, Marchioni DM. Using Two Different Approaches to Assess Dietary Patterns: Hypothesis-Driven and Data-Driven Analysis. Nutrients 2016 Sep 23;8(10):593 [FREE Full text] [doi: 10.3390/nu8100593] [Medline: 27669289]

Abbreviations

eFI: electronic frailty index
FI: frailty index
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICU: intensive care unit
HR: hazard ratio
mFI: multimorbidity frailty index
ML-mFI: machine learning multimorbidity frailty index
NHI: National Health Insurance
NHIRD: National Health Insurance Research Database

Edited by G Eysenbach; submitted 11.09.19; peer-reviewed by H Arai, CC Hsu, L Lera ; comments to author 15.10.19; revised version received 17.12.19; accepted 24.01.20; published 11.06.20.

Please cite as:
Peng LN, Hsiao FY, Lee WJ, Huang ST, Chen LK
Comparisons Between Hypothesis- and Data-Driven Approaches for Multimorbidity Frailty Index: A Machine Learning Approach
J Med Internet Res 2020;22(6):e16213
URL: http://www.jmir.org/2020/6/e16213/
doi:10.2196/16213
PMID:32525481

Original Paper

Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial

Windy SY Chan[1,2*], BPharm, MClinPharm, DHSc; Angela YM Leung[3*], BN, MHA, PhD

[1] School of Health Sciences, Caritas Institute of Higher Education, New Territories, China (Hong Kong)

[2] Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Kowloon, China (Hong Kong)

[3] Centre for Gerontological Nursing, School of Nursing, The Hong Kong Polytechnic University, Kowloon, China (Hong Kong)

[*] all authors contributed equally

Corresponding Author:
Windy SY Chan, BPharm, MClinPharm, DHSc
School of Health Sciences
Caritas Institute of Higher Education
2 Chui Ling Lane, Tseung Kwan O
New Territories
China (Hong Kong)
Phone: 852 36536664
Email: sychan@cihe.edu.hk

Related Articles:

Comment in: https://www.jmir.org/2020/10/e21505/

Comment in: https://www.jmir.org/2020/10/e24084

Abstract

Background: Social network sites (SNSs) are widely exploited in health education and communication by the general public, including patients with various conditions. Nevertheless, there is an absence of evidence evaluating SNSs in connecting health professionals for professional purposes.

Objective: This pilot randomized controlled trial was designed to evaluate the feasibility of an intervention aiming to investigate the effects of a continuous professional education program utilizing Facebook to obtain knowledge on dementia and care for patients with dementia.

Methods: Eighty health professionals from Hong Kong were recruited for participation in the study and randomized at a 1:1 ratio by a block randomization method to the intervention group (n=40) and control group (n=40). The intervention was an 8-week educational program developed to deliver updated knowledge on dementia care from a multidisciplinary perspective, either by Facebook (intervention group) or by email (control group) from October 2018 to January 2019. The primary outcomes were the effects of the intervention, measured by differences in the means of changes in pre- and postintervention scores of knowledge assessments from the 25-item Dementia Knowledge Assessment Scale (DKAS) and formative evaluation of 20 multiple choice questions. Other outcome measurements included participant compliance, participant engagement in Facebook, satisfaction, and self-perceived uses of Facebook for continuing professional education programs.

Results: Significantly more intervention group participants (n=35) completed the study than the control group (n=25) (*P*<.001). The overall retention rate was 75% (60/80). The mean of changes in scores in the intervention group were significant in all assessments (*P*<.001). A significant difference in the mean of changes in scores between the two groups was identified in the DKAS subscale Communication and Behavior (95% CI 0.4-3.3, *P*=.02). There was no significant difference in the total DKAS scores, scores of other DKAS subscales, and multiple choice questions. Participant compliance was significantly higher in the intervention group than in the control group (*P*<.001). The mean numbers of participants accessing the learning materials were 31.5 (SD 3.9) and 17.6 (SD 5.2) in the intervention and control group, respectively. Polls attracted the highest level of participant engagement, followed by videos. Intervention group participants scored significantly higher in favoring the use of Facebook for the continuing education program (*P*=.03). Overall, participants were satisfied with the interventions (mean score 4 of a total of 5, SD 0.6).

Conclusions: The significantly higher retention rate, together with the high levels of participant compliance and engagement, demonstrate that Facebook is a promising tool for professional education. Education delivered through Facebook was significantly more effective at improving participants' knowledge of how people with dementia communicate and behave. Participants demonstrated positive attitudes toward utilizing Facebook for professional learning. These findings provide evidence for the feasibility of using Facebook as an intervention delivery tool in a manner that can be rolled out into practical settings.

(J Med Internet Res 2020;22(6):e16772) doi:10.2196/16772

KEYWORDS

dementia; Facebook; social network sites; continuous professional education

Introduction

Background

Social network sites (SNSs) have become an essential part of the daily lives of billions of users globally. Facebook currently represents the largest social network in the world, which engages more than 2.45 billion users globally [1]. SNSs are widely exploited in research on health behavior-related outcomes and health communication by both the general public and patients [2,3]. A systematic review showed that health professionals have also extensively adopted SNSs for professional purposes [4]. SNSs are used to support the delivery of clinical services, make referrals, and share information [5-7], demonstrating benefits in network building and professional collaboration [8-10]. Facebook is a novel tool for enhancing educational interactions [4,11], sharing domain knowledge through developing professional communities [12,13], and staying abreast of news and information pertaining to professional needs and interests [14-16]. With increased adoption of SNSs, it is critical to evaluate the effectiveness of such utilization for professional purposes by health professionals. This pilot study is the first to evaluate the effects SNSs in continuous professional education (CPE) programs by utilizing Facebook in the delivery of education on dementia care to a cohort of health professionals.

Need for CPE on Dementia

Dementia has been acknowledged as a significant global public health issue [17,18]. In 2015, it was estimated that 46.8 million people were living with dementia worldwide, with those numbers expected to almost double every 20 years, reaching up to 131.5 million by 2050 [19]. Timely detection and proper management of dementia are critical to allow patients to plan for their future care while they still have the capacity to make important decisions [20]. However, inadequate effort has been made to address the continuous educational needs of the health care workforce on the topic of dementia [21]. Low levels of dementia knowledge among health care personnel remain common [21], representing a major barrier to providing appropriate end-of-life care to people with dementia [22]. The limited teaching hours in preregistration programs was identified as one of the main factors associated with the lack of knowledge of dementia [23]. Postregistration training on dementia care has been offered in some programs, but the most common approach was traditional small to large group face-to-face delivery [21], which has limitations in terms of resources, time, and geographical factors. Electronic learning (eLearning) options for CPE programs offer flexibility with respect to both the time and place of learning. Nevertheless, these types of programs can be accompanied by inadequate levels of participant interaction, reflection, practice, and application to practice [24,25]. With the evolving utilization of SNSs by health professionals, there is potential to make the most of SNSs to meet professional education requirements and promote interaction while avoiding a significant resource burden.

Although some research has been conducted to investigate the uses of SNSs in health communication and education among the general public and patients, there is a general lack of literature on the topic, with a notable sparsity of evidence from intervention trials on evaluating SNSs in connecting health professionals for professional communication [4]. The findings obtained to date are in line with reviews on professional dementia education, in which most of the included studies used weak research designs that were unable to identify the direct effects of the interventions [26,27].

Aim and Objectives

This pilot trial was designed to fill the gaps in the current literature, representing the first pilot randomized controlled trial (RCT) examining the effects of an SNS on professional education by health professionals. The aim of the study was to evaluate the feasibility of an interventional trial designed to investigate the effects of a CPE program utilizing Facebook on dementia knowledge and care. The objectives to achieve the research aim were to: (1) quantitatively evaluate the effects of a Facebook intervention by measuring the change in knowledge pre and postintervention, (2) quantitatively evaluate the feasibility and acceptability of the intervention, and (3) qualitatively assess participant perceptions of the intervention related to recruitment, delivery, and usability.

Methods

Overview

The study was a pilot randomized, unblinded, controlled trial with mixed-research methods (quantitative and qualitative). The intervention was an 8-week CPE program on dementia delivered to participants either by Facebook (intervention group) or by email communications (control group) from October 2018 to January 2019. Ethical approvals were obtained from the Hong Kong Polytechnic University (reference number: HSEARS20180307007) and the Caritas Institute of Higher Education (IRG180100). Trial reporting follows the CONsolidated Standards of Reporting Trials (CONSORT) 2010 statement: extension to randomized pilot and feasibility trials [28].

Subject Recruitment and Randomization

Health professionals were recruited via promotion of the trial at professional conferences, and by invitations disseminated through professional bodies and health institutes. Inclusion criteria were being a health professional, having access to the internet, in addition to having an email account and Facebook account or willing to create one otherwise. Individuals were excluded if they were working at institutes specializing in the delivery of dementia or geriatric care or had prior training in dementia care that was completed in the past 6 months.

Before randomization, participants were required to complete baseline assessments, knowledge tests, and consent forms. In addition to collecting background information such as age, gender, occupation, and years of working experience, the baseline assessment also surveyed the participants' habits with respect to the use of SNSs, including SNS choices and frequency of use.

Consenting participants were randomized to two groups at a 1:1 ratio by a block randomization method in view of the small sample size [29]. The randomization was performed by an independent researcher who had no involvement in the trial. The block size was blinded from the investigator performing this study.

Intervention

The learning materials (ie, the educational intervention) were developed by a team of experts from the relevant disciplines, including medicine, nursing, pharmacy, occupational therapy, speech therapy, and social sciences, to provide updated interdisciplinary knowledge on dementia care. Upon completion, participants would be able to (1) describe the prevalence, pathophysiology, and recognition of dementia; (2) apply the key screening and assessment tools for diagnosing dementia; (3) explain management and support to both elderly patients with dementia and their caregivers; and (4) appraise their roles in the provision of dementia care.

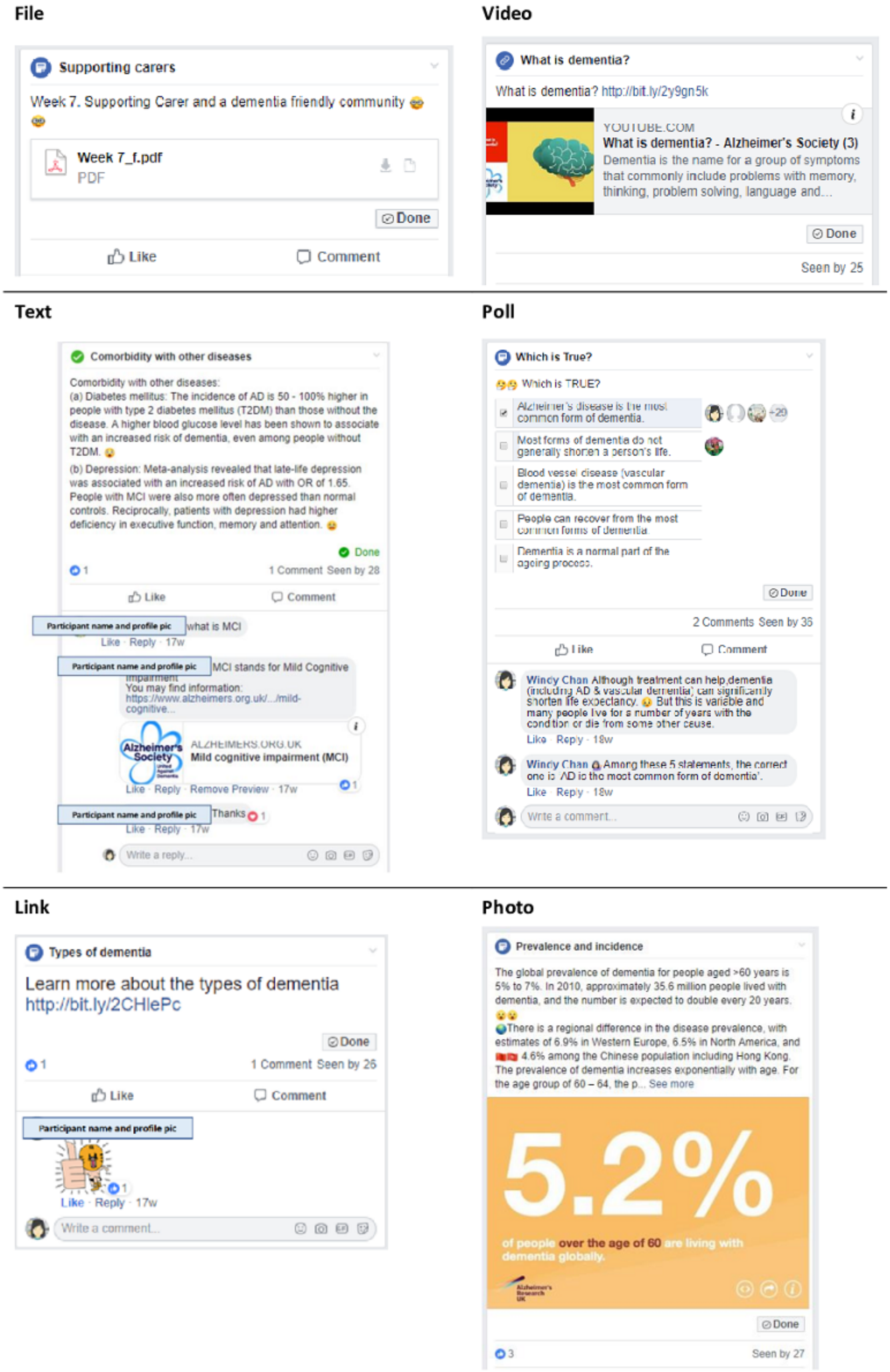
Along with the textual content, a collection of multimedia materials was selected to supplement participant learning, including websites, educational videos, pictures, diagrams, and interactive online games.

Participants in both groups received the same content and were expected to spend approximately 15 minutes daily reading the materials. Together with the time spent in assessments, the total learning hours amounted to approximately 12 hours throughout the 8-week intervention period.

In the intervention group, a Facebook Social Learning Group was created to deliver the CPE program. A Social Learning Group has additional features to an ordinary Facebook group, which facilitate the retrieval of learning materials and tracking participant engagement. The learning materials were uploaded every 1 to 3 days in the form of Facebook posts, photos, videos, text, files, or polls (Figure 1). A variety of components and post lengths were utilized, with the aim of maximizing participant engagement and interest [30,31]. There were typically 14 posts per week, including a post containing the notes (learning materials) of the week in PDF format. Participants were encouraged to comment, react, and interact with the other group members but were not allowed to create a new post.

In the control group, before the beginning of each study week, participants received emails containing their individual Bitly links to the PDF notes of that week. Bitly is a link management platform that enables analysis of the activity of each link [32]. Contents of notes were identical to those received by the intervention group participants, including the revision questions, photos, videos, games, and websites. Control group participants could access the multimedia content using the Bitly links inserted in the PDF notes, but could not take part in the polls or view the previews of links generated by Facebook. During the intervention period, control group participants received a total of 8 sets of notes.

Figure 1. Representative screenshots of six types of Facebook posts.



File

Supporting carers

Week 7. Supporting Carer and a dementia friendly community

Week 7_f.pdf
PDF

Video

What is dementia?

What is dementia? http://bit.ly/2y9gn5k

YOUTUBE.COM
What is dementia? - Alzheimer's Society (3)
Dementia is the name for a group of symptoms that commonly include problems with memory, thinking, problem solving, language and...

Seen by 25

Text

Comorbidity with other diseases

Comorbidity with other diseases:
(a) Diabetes mellitus: The incidence of AD is 50 - 100% higher in people with type 2 diabetes mellitus (T2DM) than those without the disease. A higher blood glucose level has been shown to associate with an increased risk of dementia, even among people without T2DM.
(b) Depression: Meta-analysis revealed that late-life depression was associated with an increased risk of AD with OR of 1.65. People with MCI were also more often depressed than normal controls. Reciprocally, patients with depression had higher deficiency in executive function, memory and attention.

1 · 1 Comment Seen by 28

Participant name and profile pic: what is MCI
Like · Reply · 17w

Participant name and profile pic: MCI stands for Mild Cognitive Impairment
You may find information: https://www.alzheimers.org.uk/.../mild-cognitive...
ALZHEIMERS.ORG.UK
Mild cognitive impairment (MCI)
Like · Reply · Remove Preview · 17w

Participant name and profile pic: Thanks
Like · Reply · 17w

Write a reply...

Poll

Which is True?

Which is TRUE?

☑ Alzheimer's disease is the most common form of dementia. +29
☐ Most forms of dementia do not generally shorten a person's life.
☐ Blood vessel disease (vascular dementia) is the most common form of dementia.
☐ People can recover from the most common forms of dementia.
☐ Dementia is a normal part of the ageing process.

2 Comments Seen by 36

Windy Chan Although treatment can help,dementia (including AD & vascular dementia) can significantly shorten life expectancy. But this is variable and many people live for a number of years with the condition or die from some other cause.
Like · Reply · 18w

Windy Chan Among these 5 statements, the correct one is 'AD is the most common form of dementia'.
Like · Reply · 18w

Write a comment...

Link

Types of dementia

Learn more about the types of dementia
http://bit.ly/2CHIePc

1 · 1 Comment Seen by 26

Participant name and profile pic
Like · Reply · 17w

Write a comment...

Photo

Prevalence and incidence

The global prevalence of dementia for people aged >60 years is 5% to 7%. In 2010, approximately 35.6 million people lived with dementia, and the number is expected to double every 20 years.
There is a regional difference in the disease prevalence, with estimates of 6.9% in Western Europe, 6.5% in North America, and 4.6% among the Chinese population including Hong Kong. The prevalence of dementia increases exponentially with age. For the age group of 60 – 64, the p... See more

3 · Seen by 27

Outcome Measures

Primary Outcome: Quantitative Evaluation of the Effects of a Facebook Intervention

The effectiveness of using Facebook to deliver the CPE program was measured by the knowledge gain between the pre and postintervention knowledge tests, calculated as the difference in scores. The knowledge gain in the two groups was also compared.

The knowledge tests consisted of formative evaluation, 20 multiple choice questions, and the 25-item Dementia Knowledge Assessment Scale (DKAS). Content for the multiple choice questions was drawn directly from the learning materials throughout the 8 learning weeks. The DKAS has been validated as a reliable measure of dementia knowledge for a wide spectrum of populations, including health professionals. The DKAS elucidates respondents' knowledge characteristics across four coherent domains: Causes and Characteristics, Communication and Behavior, Care Considerations, and Risks and Health Promotion [33]. A pilot assessment of the knowledge tests was performed to determine the coverage of questions, level of difficulty, and potential ceiling effects.

Quantitative Evaluation of Participant Compliance

Participant compliance was quantified by the number of week(s) (out of 8) the material was accessed by the participants, recorded by the "Seen by" feature of Facebook in the intervention group and by the Bitly link in the control group.

In Facebook, the "Seen by" feature of a post automatically records the individual group members that had seen the post (Figure 1). In the control group, each participant had their own 8 unique Bitly links to access the corresponding 8 weeks of materials. A participant clicking on the Bitly link was counted as having accessed the materials for that week. Bitly links were also created to track participant compliance in accessing the recommended external learning resources such as videos and websites.

Quantitative Evaluation of Intervention Group Participant Engagement

In studying the communication artifacts occurring in Facebook, engagement was defined as the metric of actions that participants took in the group during the intervention period. In addition to ordinary actions such as comment and reactions, a unique feature of the Facebook Social Learning Group is the "Done" button (Table 1). Participants were encouraged to click the "Done" button to share their learning progress. A set of descriptive data was also extracted from analyzing participant engagement in the intervention group, considering features that have been generally used in previous Facebook-based studies [30,31].

Table 1. Participant engagement actions on Facebook.

Engagement Action	Definition
Comment	A participant commented on a post or a comment made by another member in the group
React	A participant clicked the reaction button on a post to share different reactions: Like, Love, Haha, Wow, Sad, or Angry
Done	A participant clicked the "Done" button of a post. The "Done" button will turn green after being clicked.

Quantitative Evaluation of Participant Satisfaction

All participants were invited to complete a satisfaction evaluation form developed for the effective appraisal of training delivery for health professionals [34]. The form contained 10 5-point Likert-style positively phrased statements focusing on the achievement of program objectives, impact on learning, learning experiences, and application to clinical practice. There were 3 questions that evaluated the program length, information quantity, and difficulty level. A pilot assessment of the evaluation form was performed to determine the coverage of questions.

Quantitative Evaluation of Participants' Attitudes Toward Using Facebook for the CPE Program

Participants' attitudes on the professional use of Facebook were evaluated in a questionnaire modified from the Continuous Medical Education App Attitudes Survey Instrument, a validated measure of participant attitudes toward mobile app use in a continuous medical education program [35]. There were 10 5-point Likert-style positively phrased statements evaluating the educational value, acceptability, usability, and future feasibility of using Facebook to deliver the program. A pilot assessment of the survey was performed to determine the question coverage.

Qualitative Assessment of Feasibility and Acceptability

To investigate the pilot study's feasibility and acceptability, all participants were invited to complete a feedback form on their learning experiences and to take part in short semistructured interviews. To encourage participation of interviews, reminders were sent twice by Facebook messages (intervention group) and emails (control group).

In the feedback form, qualitative data were obtained via two free-form questions that asked the participants about their perceptions of the most valuable sections as well as suggestions for improvement.

The semistructured interviews were conducted following a flexible interview guide to explore a wider range of respondents' perceptions and comments. Participants were invited to share their views on (1) promotion, recruitment, and randomization arrangement; (2) delivery of the educational intervention; (3) usability of the learning materials; (4) feasibility and acceptability of the study; and (5) comments for further improvement. Respondents could refuse to answer any question or ask questions. Interviews were audio-recorded and transcribed manually. Data were organized using a Microsoft Excel 2016 spreadsheet.

Thematic analyses of the contents of feedback forms and semistructured interviews were conducted independently by the authors. Discrepancies on data analyses, coding, and themes generated were resolved by discussion until consensus was reached [36].

Statistical Analysis

A literature search did not identify any RCTs conducted on the effects of an educational intervention utilizing SNSs that could be used as a basis for calculating sample size. This study conservatively assumed a small effect size of 0.2. For a full RCT designed with 90% power and a two-sided 5% significance threshold, the recommended sample size for a pilot trial is 30 per group [37]. Assuming a dropout rate of 10%, the total target number of participants was 66. Slightly more participants (N=80) were recruited for this study due to positive responses from professional bodies.

Participant demographics and SNS usage characteristics were treated as categorical variables. The Chi square test and Fisher exact test were used to analyze dichotomous and categorical variables. Participant performance on the knowledge test is presented descriptively. To compare the knowledge gains of the two groups (primary outcome), an independent sample *t* test was used. The effect size of the outcome measure between the two groups was estimated according to the formula proposed by Morris [38] for mean pre-post changes between the intervention and control groups. An intention-to-treat approach was adopted in this study. Scores in the evaluation and survey on the use of Facebook (secondary outcome) are presented as descriptive statistics and were analyzed by a two-sample *t* test or Fisher exact test. *P*≤.05 was considered significant. Statistical analyses were conducted using R v. 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria) and IBM SPSS Statistics version 23 (IBM Corp, Armonk, NY, USA).

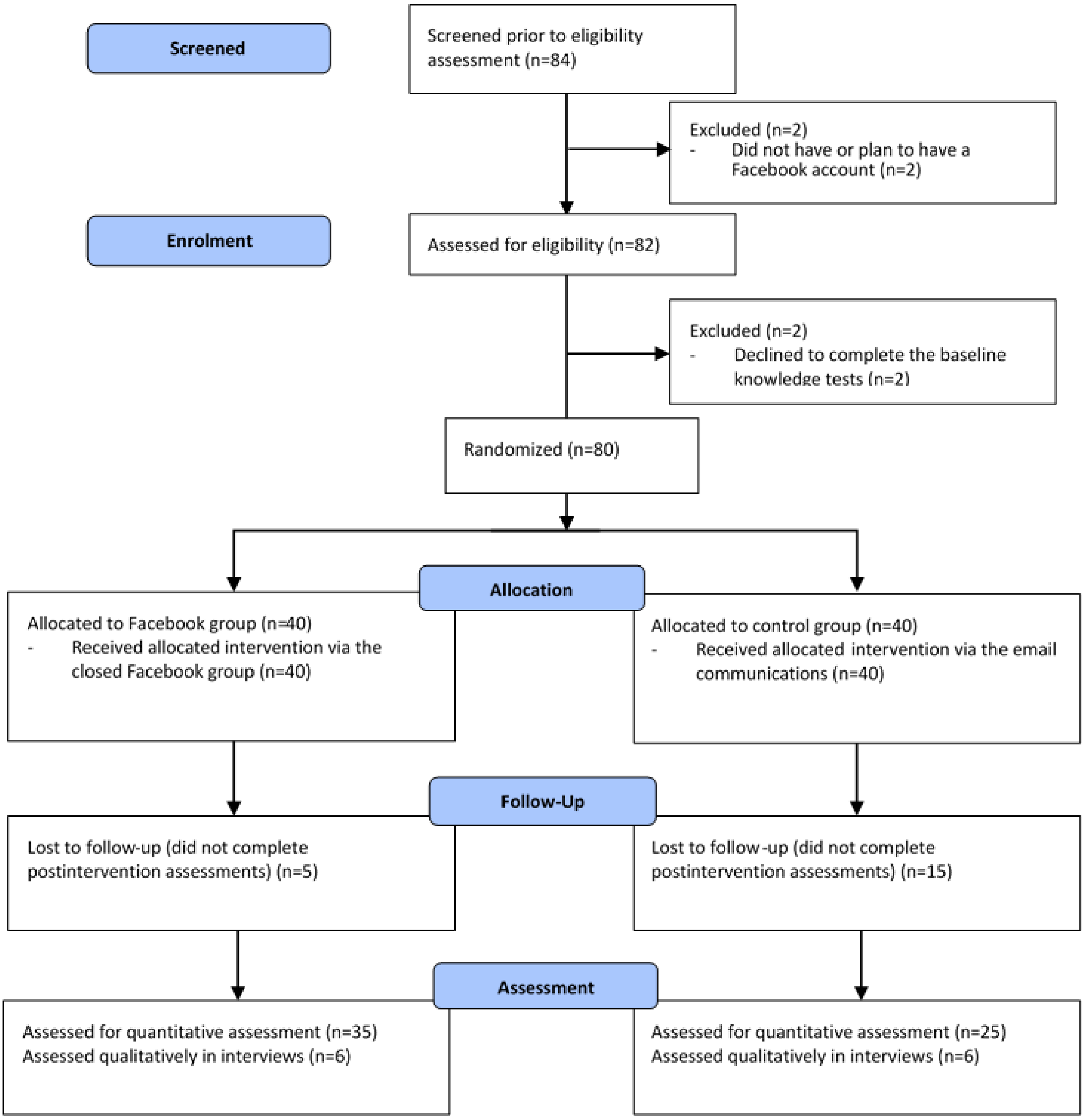
Results

Participants

Figure 2 shows the flow of study subject recruitment and randomization according to the CONSORT guidelines. Eighty participants underwent randomization with 40 allocated to the intervention group and 40 allocated to the control group.

The overall retention rate was 75% (60/80). Significantly more intervention group participants (35/40, 88%) completed the study compared to those in the control group (25/40, 63%; *P*<.001). Twelve participants, six from each group, attended postintervention interviews.

Figure 2. Study CONsolidated Standards of Reporting Trials (CONSORT) flowchart.



At baseline, participant characteristics and habits when using SNSs were comparable in the two groups (Table 2 and Multimedia Appendix 1). The majority (64%, 51/80) of participants ranged in age from 20-39 years. The participants, who worked in various health disciplines, included pharmacists, nurses, optometrists, and physiotherapists.

Among all participants, 95% were daily SNS users. WhatsApp Messenger and Facebook were the most frequently used SNSs, followed by Instagram, WeChat, and Line. SNSs were mainly used for personal use and communication as well as for general knowledge. Less than half of the participants had used SNSs for professional communication. Only two participants had previously participated in a CPE program delivered via SNSs (Multimedia Appendix 1).

At baseline, the knowledge test overall mean scores were similar in the two groups (Multimedia Appendix 2). In both groups, the lowest mean correct responses in the DKAS were found in the subscale Communication and Behavior.

Table 2. Baseline characteristics of the study participants.

Characteristic	All participants (N=80)	Intervention group (n=40)	Control group (n=40)	*P* value
Age (years), n (%)				**.80**[a]
20-29	28 (35)	14 (35)	14 (35)	
30-39	23 (29)	13 (33)	10 (25)	
40-49	15 (19)	6 (15)	9 (23)	
50-59	11 (14)	5 (13)	6 (15)	
60-69	3 (4)	2 (5)	1 (3)	
Sex				**.60**[b]
Male	19 (24)	8 (20)	11 (28)	
Female	61 (76)	32 (80)	29 (73)	
Profession				**.70**[a]
Pharmacist	33 (41)	17 (43)	16 (40)	
Nurse	30 (38)	16 (40)	14 (35)	
Others	17 (21)	7 (18)	10 (25)	
Optometrist	10 (13)	6 (15)	4 (10)	
Physiotherapist	3 (4)	1 (3)	2 (5)	
Houseman	1 (1)	0 (0)	1 (3)	
Occupational therapist	1 (1)	0 (0)	1 (3)	
Overseas pharmacist and nurse	1 (1)	0 (0)	1 (3)	
Physician	1 (1)	0 (0)	1 (3)	
Working experience				**.67**[a]
<1 year	3 (4)	0 (0)	3 (8)	
1-2 years	9 (11)	6 (15)	3 (8)	
2-3 years	3 (4)	3 (8)	0 (0)	
3-5 years	10 (13)	4 (10)	6 (15)	
5-10 years	12 (15)	7 (18)	5 (13)	
>10 years	43 (54)	20 (50)	23 (58)	

[a] Chi square test.

[b] Fisher exact test.

Quantitative Evaluation of the Effects of a Facebook Intervention

At the postintervention assessment, intervention group participants demonstrated significant knowledge gains, reflected by significant increases in the total DKAS scores, each of the DKAS subscale scores, and multiple choice questions (Table 3). In the control group, significant knowledge gains were not identified in the DKAS subscales Communication and Behavior and Care Considerations.

A significant difference in changes in the mean knowledge gain scores between the two groups was identified in the DKAS subscale Communication and Behavior, with a medium effect size (Table 3). The CPE program delivered by Facebook was more effective than the control (email provision) at increasing participants' understanding and knowledge of the communication and behavior of people with dementia. There was no significant difference between the two groups in the total DKAS, multiple choice questions, and other DKAS subscale scores.

The effect size in the change in the overall DKAS score was small. For a full-scale two-armed RCT, assuming a standard error of 5% with 95% power, a total of 272 participants would be required (ie, 136 per group). With attrition rates of 10%, 20%, or 30%, the total number of participants required is 304, 340, and 390, respectively.

Table 3. Scores and the comparison of changes in scores at postintervention knowledge assessment.

Assessment	Intervention group (n=40)			Control group (n=40)			95% CI	*P* value	Effect size
	Postintervention score, mean (SD)	Change in score, mean (SD)	*P* value	Postintervention score, mean (SD)	Change in score, mean (SD)	*P* value			
Total DKAS[a] score (maximum score 50)	36.9 (8.0)	8.3 (8.6)	<.001	35.1 (8.0)	4.9 (8.4)	.007	–1.1-7.8	.13	0.44 (small)
DKAS subscale Causes and Characteristics (maximum score 14)	11.3 (3.0)	2.5 (3.4)	<.001	10.8 (2.5)	1.8 (3.4)	.01	–1.1-2.5	.44	0.20 (small)
DKAS subscale Communication and Behavior (maximum score 12)	7.5 (3.1)	2.1 (3.6)	.001	6.2 (2.6)	0.3 (2.1)	.51	0.4-3.3	.02	0.77 (medium)
DKAS subscale Care Considerations (maximum score 12)	9.5 (2.6)	1.4 (2.8)	.007	9.9 (2.6)	1.1 (2.9)	.07	–1.2-1.8	.70	0.23 (small)
DKAS subscale Risks and Health Promotion (maximum score 12)	8.6 (2.3)	2.3 (3.0)	<.001	8.2 (2.7)	1.7 (3.0)	.009	–1.0-2.2	.48	0.21 (small)
Multiple choice question (maximum score 20)	13.2 (3.9)	4.9 (3.7)	<.001	12.8 (5.0)	3.6 (4.5)	.001	–0.9-3.5	.25	0.26 (small)

[a] DKAS: Dementia Knowledge Assessment Scale.

Quantitative Evaluation of Participant Compliance

Participant compliance with the educational intervention was significantly higher in the intervention group (mean 6.30, SD 2.28 weeks) than in the control group (mean 3.53, SD 3.21 weeks) (*P*<.001). Overall, 75% of the intervention group participants completed at least 6 weeks or more of the intervention compared to only 35% of the control group participants (Table 4). The mean number of participants accessing the materials was 31.5 (SD 3.9) and 17.6 (SD 5.2) in the intervention and control group, respectively.

Participant compliance was significantly higher in the intervention group (mean 14.9, SD 14.1) than in the control group (mean 2.1, SD 1.8) (*P*<.001) with respect to accessing the 18 external learning resources (9 videos and 9 website links) (Multimedia Appendix 3). Participants who learned via Facebook were significantly more compliant in viewing multimedia content.

Table 4. Participant compliance.

Total number of weeks accessed	Intervention group (N=40), n (%)	Control group (N=40), n (%)
0	1 (3)	9 (23)
1	2 (5)	9 (23)
2	1 (3)	1 (3)
3	2 (5)	4 (10)
4	1 (3)	2 (5)
5	3 (8)	1 (3)
6	5 (13)	1 (3)
7	7 (18)	5 (13)
8	18 (45)	8 (20)

Quantitative Evaluation of Participant Engagement on Facebook

A total of 97 Facebook posts were created for the intervention group. All posts received a considerable number of participant engagement actions (mean 19.03, range 11-33). Polls, which were question-related posts designed to encourage member interaction, attracted the highest degree of participant engagement (mean 20.17), followed by videos (mean 19.56). Links (mean 17.33) generated the lowest number of engagement actions during the intervention period.

Quantitative Evaluation of Participant Satisfaction With the Educational Intervention

As illustrated by item 10 in the satisfaction evaluation (Table 5), participants were satisfied with the training overall. There was no significant difference between the two groups. Evaluating achievement in each of the four course objectives (items 1-4), the overall mean scores ranged from 3.8 to 3.9 (SD 0.6-0.7). More than 80% of participants regarded the program length, quality, and difficulty level as appropriate.

Table 5. Participant satisfaction with the educational intervention.

Item	All (N=60)	Intervention group (n=35)	Control group (n=25)	*P* value
Achievement of objectives, mean (SD)				
1. Describe the prevalence, pathophysiology and recognition of dementia	3.9 (0.6)	3.8 (0.6)	4.0 (0.6)	.28[a]
2. Apply the key screening and assessment tools for diagnosing dementia	3.8 (0.7)	3.7 (0.7)	3.9 (0.8)	.30[a]
3. Explain the management and support to both elderly people with dementia and their caregivers	3.9 (0.6)	3.9 (0.6)	4.0 (0.6)	.26[a]
4. Appraise your role in the delivery of dementia care	3.8 (0.7)	3.7 (0.7)	3.9 (0.7)	.22[a]
Overall (sum of items 1 to 4)/20	15.4 (2.3)	15.1 (2.3)	15.8 (2.2)	.19[a]
Learning experiences, mean (SD)				
5. The training enhanced my knowledge of dementia care	4.2 (0.6)	4.2 (0.6)	4.1 (0.6)	.75[a]
6. The learning material was well organized	4.1 (0.7)	3.9 (0.8)	4.2 (0.6)	.18[a]
7. The learning material was easy to use	3.9 (0.9)	3.8 (1.0)	4.1 (0.8)	.21[a]
8. I will be able to apply the training information to my professional practice	3.8 (0.8)	3.8 (0.8)	3.8 (0.7)	.89[a]
9. I would recommend this training to others	4.0 (0.7)	3.9 (0.7)	4.0 (0.7)	.49[a]
10. Overall, I am satisfied with this training	4.0 (0.6)	4.0 (0.6)	4.0 (0.6)	.81[a]
Overall (sum of items 1 to 6)/30	23.9 (3.8)	23.6 (4.0)	24.3 (3.5)	.49[a]
Length of training, n (%)				**.71[b]**
About right	49 (82)	27 (77)	22 (88)	
Too long	10 (17)	7 (20)	3 (12)	
Too short	1 (2)	1 (3)	0 (0)	
Quality of information, n (%)				**.34[b]**
About right	51 (85)	30 (86)	21 (84)	
Not enough	4 (7)	1 (3)	3 (12)	
Too much	5 (8)	4 (11)	1 (4)	
Level of difficulty, n (%)				**.45[b]**
About right	52 (87)	29 (83)	23 (92)	
Too difficult	8 (13)	6 (17)	2 (8)	
Too easy	0 (0)	0 (0)	0 (0)	

[a] Independent sample *t* test.

[b] Chi square test.

Quantitative Evaluation of Participant Attitudes Toward Using Facebook for the CPE Program

Overall, the respondents were positive about using Facebook for a CPE program. Intervention group participants rated the use of Facebook significantly higher (*P*=.03) (Table 6).

Significant differences between the two groups were identified in responses to the survey questions "I would recommend Facebook for other CPE courses" (*P*=.01), "Using Facebook improved my learning experience" (*P*=.02), and "I would be more likely to attend a CPE/CNE course if it used Facebook" (*P*=.02).

At completion of this pilot trial, participants who had learned via Facebook were more willing to use Facebook in future CPE programs.

Table 6. Attitudes toward the use of Facebook for continuous professional education programs.

Item	All (N=60), mean (SD)	Intervention group (n=35), mean (SD)	Control group (n=25), mean (SD)	*P* value[a]
1. Using Facebook improved my learning experience	3.6 (0.8)	3.7 (0.9)	3.3 (0.7)	.02
2. Using Facebook helped me stay more engaged	3.5 (0.9)	3.6 (0.9)	3.2 (0.7)	.09
3. Using Facebook enabled me to gain more knowledge	3.4 (0.8)	3.5 (0.9)	3.3 (0.7)	.26
4. Using Facebook will help me apply what I have learned to clinical practice	3.2 (0.9)	3.4 (1.0)	3.0 (0.8)	.09
5. Using Facebook enhanced my education	3.4 (0.9)	3.5 (0.9)	3.2 (0.8)	.28
6. I would be more likely to attend a CPE[b]/CNE[c] course if it uses Facebook	3.4 (0.9)	3.6 (0.8)	3.1 (0.9)	.02
7. I am likely to use Facebook after the course is over	3.4 (0.9)	3.6 (0.9)	3.1 (1.0)	.05
8. Facebook is easy to use	3.6 (0.8)	3.8 (0.7)	3.4 (0.8)	.05
9. Facebook is intuitive to use	3.5 (0.8)	3.6 (0.8)	3.2 (0.8)	.05
10. I would recommend Facebook for other CPE courses	3.4 (1.0)	3.7 (0.9)	3.0 (0.9)	.01
Overall (sum of items 1 to 10)/50	34.3 (7.5)	36.0 (7.4)	31.8 (6.9)	.03

[a] Independent sample *t* test.

[b] CPE: continuing professional education.

[c] CNE: continuing nursing education.

Qualitative Assessment of Feasibility and Acceptability

Twenty-nine responses were collected from the feedback forms. Twelve participants, six from the intervention group and six from the control group, attended the semistructured interviews. A wide range of inputs were collected, including comments on current models of dementia care and CPE frameworks. We here focus on the findings that helped in evaluating the design of the study and the feasibility and acceptability of interventions. The qualitative data analyses led to the generation of the following themes.

Flexible and Effective Learning via Facebook

Learning through Facebook is highly mobile and flexible, without geographic or time constraints. Most of our participants easily accessed Facebook via the Facebook app installed on their mobile phones. Many stated that they seldom set aside a particular time to study, but rather tended to read posts whenever they had a chance. This was exemplified in the following participant comments: "I am glad that I can learn new knowledge during my daily commute," and "I appreciate the flexibility provided by Facebook, so I can learn whenever and wherever I want."

Furthermore, participants appreciated the conciseness and effectiveness of the education. One participant stated, "It's nice the Facebook posts are kept short, because I can complete them easily." Facebook provides a user-friendly platform that facilitates information retrieval, particularly during revision. This advantage was exemplified in participant comments such as "I hadn't expected that posts in Facebook could be organized in such a systematic way," and "I like the tracking feature in the Facebook group that clearly shows which posts I haven't read."

A majority of respondents, including a few from the control group, commented that "More CPE programs should be delivered by Facebook." A few participants concluded, "The flexibility in learning and attractiveness of the content on Facebook could not be replaced by other platforms."

Motivated and Engaged Participation on Facebook

All respondents from the intervention group appreciated the feeling of being encouraged to learn, which could be attributed to the Facebook automatic notification function. One participants stated "It's nice to receive notifications of new posts. That reminds me to study." A few participants liked to finish the new posts promptly after reading the Facebook notifications because it was similar to their usual Facebook habits.

Respondents also valued the sense of community that exists in a CPE course delivered using Facebook. In addition to the learning curriculum designed with a multidisciplinary approach, Facebook participants were also aware of their peers from different health disciplines and they appreciated the opportunity to know how others work in delivering dementia care.

The motivating interactions in the Facebook group also enhanced self-learning. Facebook participants claimed that they did notice how other members react to the posts. Some of them stated, “The polls remind me to take action, particularly when many members have voted.”

Enriched Learning With Multimedia Components

The multimedia components in the learning curriculum, such as videos and photos, were well accepted by all respondents, particularly those in the intervention group. They considered that the use of multimedia helped to more comprehensively explain concepts, which increased their motivation to learn. Some participants particularly prefer watching videos over other multimedia sources and considered videos as a good source for deepening their understanding on topics. This was exemplified in the following participant comments: “There should be more videos,” and “Some videos are quite interesting and informative, so I shared them with my friends.”

Participants found that Facebook highly facilitated the viewing of multimedia, which encouraged them to access the materials. Some of them claimed, “I watched most of the videos because they were easily played on the Facebook app.” However, the control group participants might not have utilized the multimedia to the same extent as their intervention group counterparts as described in the following section.

Low Compliance in the Control Group

In this study, control group participants generally experienced a less satisfactory learning experience and compliance. Although they might have accessed the materials, the majority did not read them all the way through. All control group respondents agreed that it was straightforward to access the learning materials, and felt that the topics and length were appropriate. Nevertheless, a majority claimed that they were not motivated to study in this one-way eLearning mode. One of the perceived obstacles was time. Control group participants typically felt that they ought to set aside a particular period of time to read through the learning materials. However, in their busy lives they usually failed to do so, and then gave up. This was exemplified by participant comments such as “The topics are interesting, but I feel that I need to set aside a particular time to read the notes. Then, I am too occupied to do so,” and “It seems too heavy and boring to read through all the notes.” Moreover, control group participants tended not to view the multimedia components, as they perceived it to be an inconvenience.

Other than receiving emails from the course administrator, control group participants did not have the chance to interact with other participants. A majority of control group respondents claimed they were not engaged in the course until the moment they were asked to complete the postintervention assessments. A few respondents commented, “I did not feel that I had been enrolled in a course until I was reminded to do the tests.”

Technical Issues Encountered in Online Learning

All respondents denied that any particular technical issues hindered their learning.

In both groups, several participants stated a concern that viewing videos on mobile phones would consume too much data. Consequently, they tended to study only when they had a Wi-Fi connection. There were suggestions to enlarge some diagrams and photos to enhance reading on a mobile phone. A few intervention group participants commented that the posts were not easy to read (the font size was quite small) on their mobile phone, and therefore they needed to use a computer.

Discussion

Principal Findings

This pilot RCT, as the first to evaluate the effects of using an SNS to deliver a CPE program, demonstrates how a novel educational intervention can result in significant knowledge gains in dementia care to health professionals. The high participant compliance and engagement, together with high satisfaction and acceptability, indicate that Facebook can be an effective and feasible learning tool that could easily be rolled out in practical settings.

Effectiveness of Facebook in Improving Knowledge

The Facebook intervention succeeded in offering the participants significant knowledge gains in all aspects of the knowledge assessments.

Comparing the means of changes in scores between the two groups, the Facebook intervention excelled at producing significant knowledge gains according to the DKAS subscale Communication and Behavior with a medium effect size. The six items included in this subscale assess how respondents interpret the behavioral symptoms of people with dementia and the proper way to communicate efficiently. Communication is a crucial skill in dementia care, but has been difficult for health personnel to learn [39]. For effective learning, rather than using plain text, all participants were instructed to view a collection of videos, websites, and online resources illustrating actual patient behaviors and communication. Although participants in both groups could access the same resources, significantly more intervention group participants accessed the materials. In addition to attracting participants by showing previews of multimedia contents, Facebook’s mobile app enables users to browse external links within the interface. In contrast, the control group participants tended to skip the links, because opening a separate browser appeared to disrupt their reading. Likewise, respondents in previous studies appraised the key benefit of SNSs as being efficient channels for a complex stream of information and multimedia files with no restriction in terms of location or office hours [40,41].

Facebook includes important features affecting participant performance in an online course, improving instructor effectiveness in facilitating learning [42]. On Facebook, posts were formatted in a concise and appealing way to reduce readers’ cognitive loads. With the Facebook tracking function, participants could review the contents nonsequentially and refer back to materials based on their study preferences. By contrast, in the control group, participants usually perceived reading the learning materials as a heavy workload that hindered their learning. Therefore, Facebook is a more effective learning platform by facilitating the integration of learning into routine habits of daily life.

Motivated Compliance and Engagement on Facebook

Learning through Facebook encouraged participant engagement and compliance. Indeed, participant retention in the intervention group (88%) was remarkably higher than that in the control group (63%) at study completion. Compared to other Facebook interventions for patients or the public, similarly high retention rates have been noted in studies promoting health-related behavioral changes such as running (78%) [43], physical activity (86%) [44], and weight loss (96%) [45]. The sense of presence and interactions with members in the Facebook group were appreciated by respondents in the evaluation and interviews. A higher degree of engagement could contribute to higher attendance [46], explaining the promise of learning through Facebook. However, the control group participants lacked a sense of belonging and involvement, which may explain the high dropout rate in this group.

Facebook is superior to email communication in providing a learner-centered, instructor-facilitated, more interactive platform for participants, which are demonstrated influencing factors of successful CPE programs [47]. Of note, in many jurisdictions, CPE schemes are voluntary, making it a challenge to motivate health professionals to complete learning programs [48,49]. The main deterrents to participation in a CPE include cost, resources, inability to get time off from work, household responsibilities, and lack of time [48,49]. Facebook features good usability to be an attractive CPE program by providing a learning platform unrestricted by time and geographical factors.

Analyses of intervention group activities were conducted to determine the post type generating the highest participant engagement. Polls attracted the most engagement, followed by videos. Links generated the lowest engagement rates, because they were perceived to lead to an external site, disrupting participants' reading on Facebook. These findings are consistent with previous studies in the marketing field investigating brand engagement [30,31].

Participants' Attitudes Toward Facebook for Professional Education

Participants in this study were found to have positive views toward utilizing Facebook for a CPE program.

Intervention group participants rated their learning experience as significantly improved when using Facebook. They were significantly more supportive of recommending the use of Facebook for other CPE courses. Participants appreciated the convenience, flexibility, and usability offered by Facebook that enables ready participation for health professionals living in different regions, negating geographical boundaries. Likewise, respondents in previous studies appraised the key benefit of SNSs as being efficient channels with no restrictions in terms of location or office hours [40,41].

With the extensive use of SNSs in our daily lives, these positive attitudes toward Facebook highlight the feasibility of conducting further research on SNS use by health professionals.

Study Strengths

This pilot RCT has several strengths, including the randomized controlled design, learning curriculum developed from a multidisciplinary approach, and evaluation of a novel educational intervention utilizing Facebook with both quantitative and qualitative research approaches. The utilization of validated tools allowed for reliable quantitative measures. The qualitative data analyses enabled gaining deeper understanding of experiences on the educational interventions and identification of factors that were important to the participating health professionals.

Limitations

There are limitations of this study that warrant mention. Health professionals who had been frequent SNS users were more willing to join this study, which introduced bias in evaluating users' perceptions of Facebook. A majority of participants were aged between 20 and 39 years. These factors would affect the generalizability of the results to the overall workforce.

Many control group participants dropped out of the study, which made the overall dropout rate (25%) larger than the assumption in estimating the sample size (10%). This may have introduced bias and reduced power, affecting the generalizability, validity, and reliability of the results [50]. This issue is slightly mitigated by the increase in sample size from 60 to 80 and the intention-to-treat approach. However, it is understood that the estimate of effect becomes conservative due to dilution as a result of dropouts [51]. Very few participants took part in the semistructured interviews. Therefore, these responses might not adequately reflect the experiences of all those who completed the study, and response bias was likely.

For practicality, the knowledge tests were self-administered, which reduces the ability to standardize test conditions. Although participants were reminded not to check other resources, it was unfeasible to prevent participants from seeking help from others or checking references.

Implications

This pilot RCT brings forward several implications for further research and practice.

Facebook was found to offer substantial advantages over conventional online CPE programs, being beneficial to those who prefer distance learning, while maintaining interactions with classmates. Leadership by the administrator and member engagement were identified as crucial factors in SNSs [52]. A larger trial is required to thoroughly examine the potential effectiveness of an SNS-based educational intervention as compared to traditional learning approaches. Given the effect size of 0.44 of this study, the total number of participants required would be 272 for a full-scale two-armed RCT.

The benefits of utilizing multimedia components on Facebook highlight strategies for designing suitable educational content delivered by SNSs. Facebook can be an effective training tool for understanding patients or caregiver communication and behavior. Not restricted to dementia care, Facebook may also be used to deliver training on other mental health conditions such as depression and schizophrenia, or behavioral conditions, including speech and movement disorders.

In future research, in addition to measuring changes in knowledge, the impacts on practical implementation and even patient care would be important to investigate. However, the complexity of behavioral change increases as evaluation of an intervention ascends the hierarchy [53]. For dementia, there are some feasible, validated measurement tools available for this purpose, including attitudes toward dementia and self-confidence in caring for people with dementia [54].

Further research would consider including an economic analysis to investigate the cost of delivering CPE through Facebook compared to other channels such as face-to-face meetings and email communication. One of Facebook's economic advantages is that the post content can easily be reused. Owing to Facebook's user-friendly features such as post scheduling, the resources required for program administration are minimal.

Conclusion

Significantly more intervention group participants completed the study than control group participants (*P*<.001). The high levels of participant compliance and engagement make Facebook a promising method for delivering professional educational programs. Attending a CPE program through Facebook significantly improved participants' knowledge of dementia (*P*<.001). Compared to the control group, although a significant difference was not observed in the mean of changes in total scores, the Facebook intervention was significantly more effective at improving participants' knowledge of communication and behavior in dementia patients (*P*=.02). Participants demonstrated positive attitudes toward utilizing Facebook for professional learning. As the first pilot RCT examining the effectiveness of an SNS for professional education for health professionals, the findings of this study provide evidence for the feasibility of using Facebook for intervention delivery in a practical setting.

Acknowledgments

This project was funded by the Internal Research Grant of the Caritas Institute of Higher Education (IRG180100).

Authors' Contributions

All authors made significant contributions to the planning, recruitment, execution, analysis, and publication of this study.

Conflicts of Interest

None declared.

This randomized study was not registered as it is an educational intervention involving health care professionals only. The editor granted an exception as ICMJE does not require a registration if the purpose of the study is to examine the effect only on health care providers rather than patients. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Participants' habits in using social network sites at baseline.
[DOCX File , 31 KB - jmir_v22i6e16772_app1.docx]

Multimedia Appendix 2
Scores in preintervention knowledge tests.
[DOCX File , 29 KB - jmir_v22i6e16772_app2.docx]

Multimedia Appendix 3
Participant compliance: Number of clicks per external resources.
[DOCX File , 31 KB - jmir_v22i6e16772_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 1277 KB - jmir_v22i6e16772_app4.pdf]

References

1. Statista. 2019. Most popular social networks worldwide as of July 2019, ranked by number of active users (in millions), URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2019-10-02]

2. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
3. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De BI, Thomas S, et al. A Web-Based, Social Networking Physical Activity Intervention for Insufficiently Active Adults Delivered via Facebook App: Randomized Controlled Trial. J Med Internet Res 2015;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
4. Chan WS, Leung AY. Use of Social Network Sites for Communication Among Health Professionals: Systematic Review. J Med Internet Res 2018 Mar 28;20(3):e117 [FREE Full text] [doi: 10.2196/jmir.8382] [Medline: 29592845]
5. Gulacti U, Lok U, Hatipoglu S, Polat H. An Analysis of WhatsApp Usage for Communication Between Consulting and Emergency Physicians. J Med Syst 2016 Jun;40(6):130. [doi: 10.1007/s10916-016-0483-8] [Medline: 27083574]
6. Sarode SC, Sarode GS, Anand R, Patil S, Unadkat H. WhatsApp is an effective tool for obtaining second opinion in oral pathology practice. J Oral Pathol Med 2017 Aug 10;46(7):513-519. [doi: 10.1111/jop.12515] [Medline: 27770472]
7. Johnston MJ, King D, Arora S, Behar N, Athanasiou T, Sevdalis N, et al. Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams. Am J Surg 2015 Jan;209(1):45-51. [doi: 10.1016/j.amjsurg.2014.08.030] [Medline: 25454952]
8. Dong C, Cheema M, Samarasekera D, Rajaratnam V. Using LinkedIn for Continuing Community of Practice Among Hand Surgeons Worldwide. J Contin Educ Health Prof 2015;35(3):185-191. [doi: 10.1002/chp.21300] [Medline: 26378424]
9. Fuoco M, Leveridge MJ. Early adopters or laggards? Attitudes toward and use of social media among urologists. BJU Int 2015 Mar;115(3):491-497. [doi: 10.1111/bju.12855] [Medline: 24981237]
10. Raiman L, Antbring R, Mahmood A. WhatsApp messenger as a tool to supplement medical education for medical students on clinical attachment. BMC Med Educ 2017 Jan 06;17(1):7 [FREE Full text] [doi: 10.1186/s12909-017-0855-x] [Medline: 28061777]
11. Pickering JD, Bickerdike SR. Medical student use of Facebook to support preparation for anatomy assessments. Anat Sci Educ 2017 Jun;10(3):205-214. [doi: 10.1002/ase.1663] [Medline: 27806192]
12. Rolls K, Hansen M, Jackson D, Elliott D. How Health Care Professionals Use Social Media to Create Virtual Communities: An Integrative Review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
13. Gonzalez RS, Amer SM, Yahia NB, Costa FD, Noatay M, Qiao J, et al. Facebook Discussion Groups Provide a Robust Worldwide Platform for Free Pathology Education. Arch Pathol Lab Med 2017 May;141(5):690-695. [doi: 10.5858/arpa.2016-0369-OA] [Medline: 27763789]
14. Kostka-Rokosz MD, Camiel LD, McCloskey WW. Pharmacy students' perception of the impact of a Facebook-delivered health news service—Two-year analysis. Curr Pharm Teach Learn 2014 Jul;6(4):471-477. [doi: 10.1016/j.cptl.2014.04.007]
15. Cain J, Scott DR, Tiemeier AM, Akers P, Metzger AH. Social media use by pharmacy faculty: Student friending, e-professionalism, and professional use. Curr Pharm Teach Learn 2013 Jan;5(1):2-8. [doi: 10.1016/j.cptl.2012.09.002]
16. Marteau TM, Dormandy E. Facilitating informed choice in prenatal testing: how well are we doing? Am J Med Genet 2001;106(3):185-190. [doi: 10.1002/ajmg.10006] [Medline: 11778978]
17. Larson EB, Yaffe K, Langa KM. New insights into the dementia epidemic. N Engl J Med 2013 Dec 12;369(24):2275-2277 [FREE Full text] [doi: 10.1056/NEJMp1311405] [Medline: 24283198]
18. World Health Organization. Dementia: A Public Health Priority. United Kingdom: World Health Organization; 2020.
19. Prince M, Wimo A, Guerchet M, Ali G, Wu Y, Prina M. World Alzheimer Report 2015 - The Global Impact of Dementia. London: Alzheimer's Disease International; 2015.
20. Prince M, Bryce R, Ferri C. World Alzheimer Report 2011: The benefits of early diagnosis and intervention. In: World Alzheimer Report 2011. London: Alzheimer's Disease International; 2011.
21. Surr CA, Gates C, Irving D, Oyebode J, Smith SJ, Parveen S, et al. Effective Dementia Education and Training for the Health and Social Care Workforce: A Systematic Review of the Literature. Rev Educ Res 2017 Oct 31;87(5):966-1002 [FREE Full text] [doi: 10.3102/0034654317723305] [Medline: 28989194]
22. Dening KH, Greenish W, Jones L, Mandal U, Sampson EL. Barriers to providing end-of-life care for people with dementia: a whole-system qualitative study. BMJ Support Palliat Care 2012 Jun;2(2):103-107. [doi: 10.1136/bmjspcare-2011-000178] [Medline: 24654049]
23. Kwok T, Lam K, Yip A, Ho F. Knowledge of Dementia Among Undergraduates in the Health and Social Care Professions in Hong Kong. Social Work in Mental Hlth 2011 Jul;9(4):287-301. [doi: 10.1080/15332985.2011.572696]
24. Lawn S, Zhi X, Morello A. An integrative review of e-learning in the delivery of self-management support training for health professionals. BMC Med Educ 2017 Oct 10;17(1):183 [FREE Full text] [doi: 10.1186/s12909-017-1022-0] [Medline: 29017521]
25. Khazanova D, Safdieh JE. Continuing Medical Education in Neurology. Semin Neurol 2018 Aug;38(4):479-485. [doi: 10.1055/s-0038-1666978] [Medline: 30125903]
26. Jackson M, Pelone F, Reeves S, Hassenkamp AM, Emery C, Titmarsh K, et al. Interprofessional education in the care of people diagnosed with dementia and their carers: a systematic review. BMJ Open 2016 Aug 16;6(8):e010948 [FREE Full text] [doi: 10.1136/bmjopen-2015-010948] [Medline: 27531724]

27. Rampatige R, Dunt D, Doyle C, Day S, van Dort P. The effect of continuing professional education on health care outcomes: lessons for dementia care. Int Psychogeriatr 2009 Apr;21(Suppl 1):S34-S43. [doi: 10.1017/S1041610209008746] [Medline: 19288970]
28. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
29. Kendall JM. Designing a research project: randomised controlled trials and their principles. Emerg Med J 2003 Mar;20(2):164-168 [FREE Full text] [doi: 10.1136/emj.20.2.164] [Medline: 12642531]
30. Malhotra A. MIT Sloan Management Review. 2013. How to Create Brand Engagement on Facebook URL: https://sloanreview.mit.edu/article/how-to-create-brand-engagement-on-facebook/ [accessed 2019-10-02]
31. Sabate F, Berbegal-Mirabent J, Cañabate A, Lebherz P. Factors influencing popularity of branded content in Facebook fan pages. Eur Manage J 2014 Dec;32(6):1001-1011. [doi: 10.1016/j.emj.2014.05.001]
32. Bitly. 2017. What is Bitly? URL: https://support.bitly.com/hc/en-us/articles/230895688-What-is-Bitly- [accessed 2019-10-02]
33. Annear MJ, Toye C, Elliott KJ, McInerney F, Eccleston C, Robinson A. Dementia knowledge assessment scale (DKAS): confirmatory factor analysis and comparative subscale scores among an international cohort. BMC Geriatr 2017 Jul 31;17(1):168 [FREE Full text] [doi: 10.1186/s12877-017-0552-y] [Medline: 28760154]
34. Northwest Center for Public Health Practice. Effective Evaluation: A Toolkit for Evaluating Presentations and Trainings URL: http://www.nwcphp.org/training/effective-evaluation-a-toolkit-for-evaluating-presentations-and-trainings [accessed 2019-10-02]
35. Wittich CM, Wang AT, Fiala JA, Mauck KF, Mandrekar JN, Ratelle JT, et al. Measuring Participants' Attitudes Toward Mobile Device Conference Applications in Continuing Medical Education: Validation of an Instrument. J Contin Educ Health Prof 2016;36(1):69-73. [doi: 10.1097/CEH.0000000000000031] [Medline: 26954248]
36. Wicks D. The Coding Manual for Qualitative Researchers (3rd edition)The Coding Manual for Qualitative Researchers (3rd edition) Johnny Saldaña Sage 2015 ISBN-13: 978-1473902497. Qual Research Orgs Mgmt 2017 Jun 12;12(2):169-170. [doi: 10.1108/qrom-08-2016-1408]
37. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
38. Morris SB. Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. Organ Res Meth 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]
39. Riachi R. Person-centered communication in dementia care: a qualitative study of the use of the SPECAL® method by care workers in the UK. J Soc Work Pract 2017 Oct 26;32(3):303-321. [doi: 10.1080/02650533.2017.1381948]
40. BuzzSumo. 2019. The 2019 Ultimate Guide to Facebook Engagement URL: https://buzzsumo.com/blog/facebook-engagement-guide/ [accessed 2019-10-02]
41. Winandy M, Kostkova P, de QE, St LC, Szomszor M. Follow #eHealth2011: Measuring the Role and Effectiveness of Online and Social Media in Increasing the Outreach of a Scientific Conference. J Med Internet Res 2016 Jul 19;18(7):e191 [FREE Full text] [doi: 10.2196/jmir.4480] [Medline: 27436012]
42. McFarland D, Hamilton D. Factors Affecting Student Performance and Satisfaction: Online versus Traditional Course Delivery. J Comput Inform Syst 2005;46(2):25-32. [doi: 10.1080/08874417.2006.11645880]
43. Looyestyn J, Kernot J, Boshoff K, Maher C. A Web-Based, Social Networking Beginners' Running Intervention for Adults Aged 18 to 50 Years Delivered via a Facebook Group: Randomized Controlled Trial. J Med Internet Res 2018 Feb 26;20(2):e67 [FREE Full text] [doi: 10.2196/jmir.7862] [Medline: 29483065]
44. Kernot J, Olds T, Lewis LK, Maher C. Usability testing and piloting of the Mums Step It Up program--a team-based social networking physical activity intervention for women with young children. PLoS One 2014 Oct;9(10):e108842 [FREE Full text] [doi: 10.1371/journal.pone.0108842] [Medline: 25272039]
45. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31. [doi: 10.1002/oby.20232] [Medline: 23505165]
46. Devadoss S, Foltz J. Evaluation of Factors Influencing Student Class Attendance and Performance. Am J Agricult Econ 1996 Aug 01;78(3):499-507. [doi: 10.2307/1243268]
47. Brekelmans G, Maassen S, Poell RF, Weststrate J, Geurdes E. Factors influencing nurse participation in continuing professional development activities: Survey results from the Netherlands. Nurse Educ Today 2016 May;40:13-19. [doi: 10.1016/j.nedt.2016.01.028] [Medline: 27125144]
48. Gibson JM. Using the Delphi technique to identify the content and context of nurses' continuing professional development needs. J Clin Nurs 1998 Sep;7(5):451-459. [doi: 10.1046/j.1365-2702.1998.00175.x] [Medline: 9855997]
49. Schweitzer DJ, Krassa TJ. Deterrents to nurses' participation in continuing professional development: an integrative literature review. J Contin Educ Nurs 2010 Oct;41(10):441-447; quiz 448. [doi: 10.3928/00220124-20100601-05] [Medline: 20540461]
50. Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child 2008 Jun;93(6):458-461. [doi: 10.1136/adc.2007.127316] [Medline: 18495909]

51. Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res 2011 Jul;2(3):109-112. [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
52. Gruzd A, Haythornthwaite C. Enabling community through social media. J Med Internet Res 2013;15(10):e248 [FREE Full text] [doi: 10.2196/jmir.2796] [Medline: 24176835]
53. Kirkpatrick D. Implementing the four levels: a practical guide for effective evaluation of training programs. 1st ed. San Francisco: Berrett-Koehler Publishers; 2007.
54. Travers CM, Beattie E, Martin-Khan M, Fielding E. A survey of the Queensland healthcare workforce: attitudes towards dementia care and training. BMC Geriatr 2013 Sep 30;13(1):101. [doi: 10.1186/1471-2318-13-101]

Abbreviations

CONSORT: CONsolidated Standards of Reporting Trials
CPE: Continuous professional education
DKAS: Dementia Knowledge Assessment Scale
eLearning: electronic learning
RCT: randomized controlled trial
SNS: social network site

Edited by G Eysenbach; submitted 23.10.19; peer-reviewed by S Wawrzynski, K Rolls; comments to author 23.12.19; revised version received 27.02.20; accepted 20.03.20; published 02.06.20.

Please cite as:
Chan WSY, Leung AYM
Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial
J Med Internet Res 2020;22(6):e16772
URL: https://www.jmir.org/2020/6/e16772
doi:10.2196/16772
PMID:32484441

Original Paper

Teaching the National Institutes of Health Stroke Scale to Paramedics (E-Learning vs Video): Randomized Controlled Trial

Avinash Koka[1*], MD; Laurent Suppan[1*], MD; Philippe Cottet[1], MD; Emmanuel Carrera[2], MD; Loric Stuby[3], EMT-P; Mélanie Suppan[4], MD

[1]Division of Emergency Medicine, Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals, Geneva, Switzerland

[2]Stroke Center, Department of Neurology, Geneva University Hospitals, Geneva, Switzerland

[3]Genève TEAM Ambulances, Geneva, Switzerland

[4]Division of Anesthesiology, Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
Laurent Suppan, MD
Division of Emergency Medicine
Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
Geneva University Hospitals
Rue Gabrielle-Perret-Gentil 4
Geneva, 1211
Switzerland
Phone: 41 795532579
Email: laurent.suppan@hcuge.ch

Abstract

Background: Prompt and accurate identification of stroke victims is essential to reduce time from symptom onset to adequate treatment and to improve neurological outcomes. Most neurologists evaluate the extent of neurological deficit according to the National Institutes of Health Stroke Scale (NIHSS), but the use of this scale by paramedics, the first healthcare providers to usually take care of stroke victims, has proven unreliable. This might be, at least in part, due to the teaching method. The video used to teach NIHSS lacks interactivity, while more engaging electronic learning (e-learning) methods might improve knowledge acquisition.

Objective: This study was designed to evaluate whether a highly interactive e-learning module could enhance NIHSS knowledge acquisition in paramedics.

Methods: A randomized controlled trial comparing a specially designed e-learning module with the original NIHSS video was performed with paramedics working in Geneva, Switzerland. A registration number was not required as our study does not come into the scope of the Swiss federal law on human research. The protocol was nevertheless submitted to the local ethics committee (Project ID 2017-00847), which issued a “Declaration of no objection.” Paramedics were excluded if they had prior knowledge of or previous training in the NIHSS, or if they had worked in a neurology or neurosurgery ward. The primary outcome was overall performance in the study quiz, which contained 50 questions. Secondary outcomes were performance by NIHSS item, time to course and quiz completion, user satisfaction regarding the learning method, user perception of the course duration, and probability the user would recommend the course to a colleague.

Results: The study was completed by 39 paramedics. There was a better overall median score (36/50 vs 33/50, *P*=.04) and a higher degree of satisfaction regarding the learning method in the e-learning group (90% vs 37%, *P*=.002). Users who had followed the e-learning module were more likely to recommend the course to a colleague (95% vs 63%, *P*=.02). Paramedics in the e-learning group took more time to complete the course (93 vs 59 minutes, *P*<.001), but considered the duration to be more adequate (75% vs 32%, *P*=.01). Time to quiz completion was similar between groups (25 vs 38 minutes, *P*=.12).

Conclusions: Use of an e-learning module shows promising results in teaching the NIHSS to paramedics.

(J Med Internet Res 2020;22(6):e18358) doi:10.2196/18358

KEYWORDS

active learning; electronic learning; video; stroke; online learning; e-learning

Introduction

Background and Importance

Prompt and accurate identification of stroke victims is essential to reduce time from symptom onset to adequate treatment and to improve neurological outcomes [1]. Evaluation of the extent of the neurological deficit is generally performed according to the National Institutes of Health Stroke Scale (NIHSS), which was first described in 1989 [2,3]. Although the NIHSS has been extensively studied and validated in the hospital setting for many different types of provider, data regarding its application in the prehospital setting is limited [4]. Moreover, in contrast to other stroke scales, use of the NIHSS by Helicopter Emergency Medical Services crews has proved unreliable [5-8]. Nevertheless, the NIHSS enables identification of neurological deficits often missed when using other scales, particularly in cases of cerebellar stroke [9]. Accurate and reliable use of the same scale as that used by hospital neurologists might enable paramedics to improve the prehospital triage of stroke patients, and could help decrease door-to-needle time in patients for whom thrombolysis can be considered [10].

Many different factors might contribute to the difficulty of using the NIHSS in acute stroke, including the method used to teach this scale to paramedics. The video developed by Patrick Lyden et al is the most studied method used to teach NIHSS to in-hospital providers [11]. This media has many advantages, as it does not require the presence of a teacher or of a patient, and can be used at any time, even during night shifts. Lack of interactivity, however, has also been noted as a potential shortcoming [4].

Electronic-learning (e-learning) in its various forms is increasingly used in health professional education. Highly interactive self-paced modules are part of current e-learning models and often elicit higher satisfaction in the learning population [12,13]. Greater satisfaction with the learning method is usually associated with better understanding and performance [14].

Goal of This Investigation

The aim of our study was to evaluate the performance of an e-learning module in teaching the NIHSS to paramedics. Our hypothesis was that an e-learning module should, through its higher degree of interactivity, increase the paramedics' performance.

Methods

Study Design and Setting

This was a randomized (1:1), controlled, outcome assessor and data analysts blinded superiority study compliant with the CONSORT-EHEALTH guidelines [15]. The study took place in July 2019 in Geneva, Switzerland. In this state, paramedics need to follow a 3-year curriculum before certification, during which they are taught to assess potential stroke patients using simplified scales such as the Cincinnati and G-FAST, but not the NIHSS. Geneva emergency dispatchers send an ambulance staffed with two certified paramedics whenever they identify a possible stroke. Medical reinforcement by a physician-staffed mobile unit (Service Mobile d'Urgence et de Réanimation – SMUR) is only sent if an immediate vital threat is identified by the dispatchers or ambulance crew [16].

The study protocol (Project ID 2017-00847) was submitted to the local ethics committee, which issued a "Declaration of no objection" as this project does not come into the scope of the Swiss federal law on human research. A registration number was therefore not required.

Subjects and Enrollment

All paramedics working for any of the 7 ambulance services in Geneva and/or in the SMUR unit were invited to join the study on a voluntary basis. Four training sessions were organized on two different days to ensure maximum participation despite the shift work pattern. Paramedics were excluded if they had prior knowledge of or previous training in the NIHSS, or if they had worked in a neurology or neurosurgery ward. Consent was obtained through the online platform.

Learning Material and Online Platform

An e-learning module containing 184 highly interactive slides was created by 4 of the authors (MS, LSt, LSu, and AK) under Storyline 3 (Articulate Global). The entire module is available online at https://e-learning.nihss-study.ch. To make navigation easier and to allow the end user to easily review the learning material, this module is divided into chapters, each representing one item on the NIHSS. Each chapter is composed of 3 parts: an initial description explaining the scoring of the item, an illustration of how to test the patient, and a summary of the scoring and key pitfalls to look out for. To increase interactivity and user involvement, challenging questions are presented to the participants at the beginning of each chapter prior to displaying key learning material to stimulate curiosity [17]. Each wrong answer leads to a cue or to a detailed explanation before the user can either retry or move forward. Completion of all the chapters enables access to a practice quiz covering all scoring items.

Parts of the original video by Lyden et al are embedded in the e-learning module to illustrate clinical testing, with permission of the author. The original video had to be subtitled as English language skills vary widely among the paramedics working in Geneva. It was therefore fully translated by one author (MS), who also created the required subtitles. Three of the other authors (AK, LSt, LSu) reviewed the subtitled video to ensure the quality and accuracy of the translation. The subtitled video is available at https://video.nihss-study.ch.

The learning material was uploaded to an online platform developed under the Joomla! 3.9 Content Management System (Open Source Matters).

Study Sequence

Once paramedics had registered for the study, one investigator (LSt) sent the total number of registered participants to a second investigator (MS). This second investigator, thus blinded to the participants' information, performed randomization with a 1:1 allocation using a computer-generated table (https://www.randomizer.org/), and created unique identifiers (usernames and passwords). These identifiers, that would allow paramedics to log into the e-learning platform, were printed and placed in identical, opaque, sealed envelopes. These envelopes were shuffled and randomly scattered on a table by one of the co-authors (LSt or LSu) before each study session. Paramedics were asked to randomly choose one of the envelopes upon entering the study room, and to sit in front of a computer. Paramedics discovered their allocated group (e-learning vs video) only after logging into the online platform. They were then asked a set of questions designed to gather demographic data and to screen for exclusion criteria before being able to access the learning material.

The control group watched the subtitled video with unrestricted access to the controls, allowing them to pause, rewind, or move forward as they wished. The experiment group was given access to the e-learning module and could freely review and navigate the e-learning chapters at will. Each group only had access to their learning media, and no time constraints were imposed.

All participants could take notes while consulting the learning material and use them during the study quiz but were not provided with a NIHSS form. For both groups, access to the study quiz was enabled only after completion of the learning material. The study quiz was identical for all participants and contained 50 questions. The first 5 questions were related to key NIHSS concepts and were followed by full NIHSS evaluations of three patients. These evaluations were taken from the Lyden certification videos and were displayed according to the regular NIHSS scoring logic. To allow the candidates to immediately score the item they had viewed, each evaluation was divided into 15 subtitled video extracts, which could be paused, forwarded, or rewound at will. None of these extracts were shown to the paramedics either in the control video or in the e-learning module.

Once the quiz was completed, a last set of questions, based on a 5-point Likert scale, was displayed in order to assess the paramedics' feelings and thoughts regarding the length of the course, their satisfaction regarding the teaching method, the perceived difficulty, and the probability they would recommend the course to a colleague.

Data Collection

Data was electronically and automatically recorded in a MariaDB database version 5.5.5 (MariaDB Foundation) before being extracted to an Excel spreadsheet (Microsoft Corporation). Group-related data was coded before the spreadsheet was sent for data analysis to the same investigator who performed the randomization (MS) in order to ensure the blinding.

The original data have been deposited to Mendeley Data [18].

Outcomes

The primary outcome was overall performance in the study quiz. Secondary outcomes were performance by NIHSS item, time to course and to quiz completion, user satisfaction with the learning method, user perception of the duration of the course, and probability that the user would recommend the training to a colleague.

Statistical Analysis

Based on prior teaching experience and clinical assumption, we determined that a sample size of 34 (17 per group) was required to have an 80% chance of detecting a difference of 10% in the overall performance at the 5% level of significance.

Given the low sample size, we decided to use only non-parametric statistical tests. The Fisher exact test was used for categorical variables and the Mann-Whitney U test for continuous variables. A two-sided *P*-value <.05 was considered significant.

Data analysis was performed using STATA 16.0 (Stata Corporation).

Results

Characteristics of Study Subjects

Of the 40 paramedics who registered to one of the four study sessions, 39 (98%) completed the study, none of whom met any of the exclusion criteria. One paramedic who was expected to participate did not appear for the study session. Gender distribution was the only statistically significant variable between groups (2/19 women in the e-learning group vs 10/20 in the video group, *P*=.01) (Table 1).

Table 1. Participant characteristics.

Characteristic	Video (n=19)	E-learning (n=20)	*P* value
Age range (years), n			.46
18-25	2	1	
26-30	3	7	
31-35	6	6	
36-40	4	3	
≥41	4	3	
Years since diploma, median (Q1;Q3)	5 (2;8)	6.5 (2.5;9)	.68
Gender, female, n (%)	2 (11)	10 (50)	.01

Main Results

Participants who followed the e-learning module performed better than those who followed the video—36 (34;37) vs 33 (31;38), *P*=.04. Paramedics who had followed the e-learning module had better scores regarding key NIHSS concepts (*P*=.01), the consciousness—commands item (*P*=.03), and the ataxia item (*P*=.02). Paramedics in the video group evaluated dysarthria better than those in the e-learning group (*P*=.02). Detailed results are shown in Table 2.

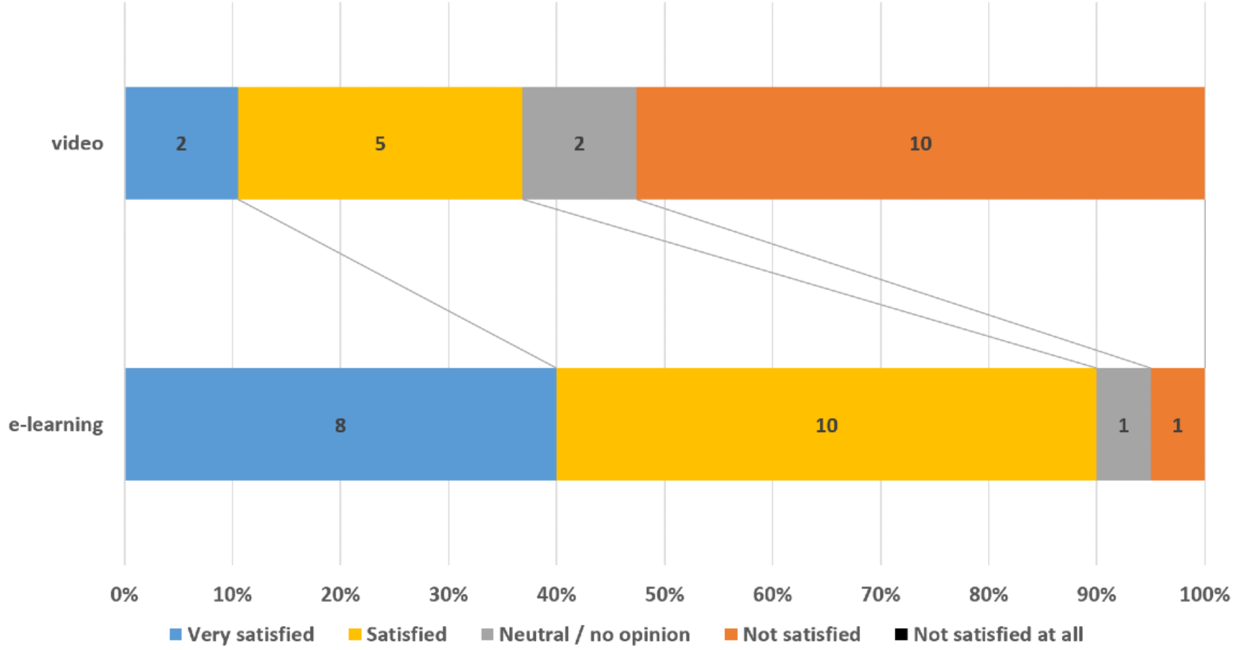
The e-learning module took more time to complete than the video with a median time of 93 (81.5;107.5) vs 59 (55;80) minutes, *P*<.001, though time to quiz completion was not significantly different between groups (*P*=.12) (Table 2). Duration of the course was considered adequate by 75% of the paramedics who had followed the e-learning module, but only by 32% of those who had followed the video (*P*=.01). Satisfaction regarding the learning method was also higher in the e-learning group compared to the video group (90% vs 37%, *P*=.002) (Figure 1). The paramedics who had followed the e-learning module were more likely to recommend the course than those who had followed the subtitled video (95% vs 63%, *P*=.02).

Table 2. Quiz results. Data are expressed as median (Q1;Q3).

Variable	Video (n=19)	E-learning (n=20)	*P* value
Overall score	33 (31;38)	36 (34;37)	.04
Time to course completion (min)	59 (55;80)	93 (81.5;107.5)	<.001
Time to quiz completion (min)	38 (33;41)	35 (31;39)	.12
Detailed results by item			
Key NIHSS[a] concepts	5 (3;5)	5 (5;5)	.01
Consciousness–Global	2 (2;2)	2 (2;3)	.30
Consciousness–Questions	3 (2;3)	3 (2;3)	.40
Consciousness–Commands	2 (2;2)	3 (2;3)	.03
Gaze	3 (2;3)	3 (2;3)	.85
Visual	2 (2;2)	2 (2;2)	1
Facial Palsy	1 (0;2)	1 (0;2)	.71
Motor arm	4 (3;5)	4 (4;5)	.75
Motor leg	5 (4;6)	5 (4;5)	.69
Ataxia	1 (0;1)	1 (1;2)	.02
Sensory	2 (2;3)	3 (2;3)	.40
Language	1 (1;1)	1 (1;2)	.41
Dysarthria	2 (2;3)	2 (2;2)	.02
Extinction and inattention	2 (1;3)	3 (2;3)	.35

[a] NIHSS: National Institutes of Health Stroke Scale.

Figure 1. Satisfaction by group.



Discussion

Main Results

The participants who followed the e-learning module performed slightly better than those who watched the subtitled version of the original NIHSS video. Our results show that paramedics who followed the e-learning module had a better understanding of the NIHSS key concepts. This difference might prove useful for the application of the NIHSS in the field.

While some specific NIHSS items, such as ataxia, were scored more accurately by the paramedics who followed the e-learning module, participants in the video group evaluated the dysarthria item significantly better. After careful review of the module, it appears that we did not include extracts from the original video for all 15 items. Indeed, the e-learning module used in this study did not contain any video extract demonstrating how to assess dysarthria. This shortcoming might, at least in part, explain why the paramedics who followed the e-learning module did not perform as well as those who had seen the video. The e-learning module should therefore be updated to include video extracts of each specific item before being tested again to assess whether our hypothesis holds true. The difference observed regarding dysarthria evaluation might also be explained by the fact that this item has been shown to have a low interrater reliability [19].

The e-learning module took longer than the video to complete, with a median difference of 34 minutes. Nevertheless, 75% of the paramedics who had followed the e-learning module judged its length adequate, whereas less than a third of those who had followed the video were of the same opinion. Three main factors might contribute to this finding. First, while the video does not allow for interactions other than pausing, forwarding, or rewinding, the e-learning module stimulates and engages the learners. Second, the possibility of quickly going back to the learning chapter one wishes to review is facilitated by the e-learning interface, as the learner loses less time looking for the right sequence. Last, the introductory chapter in the e-learning module seems to lead to an enhanced understanding of key NIHSS concepts, and might therefore help the learner better understand the goals of the course, therefore acting as a primer by increasing both the learner’s commitment and attention.

The higher level of satisfaction with the learning method is probably most strongly linked to the high interactivity of the e-learning module. Previous studies have indeed shown that a greater level of interactivity increases learner engagement while decreasing the attrition rate [20]. Though other factors such as quality of content, convenience of the technology, and quality of the support received have also been found to increase learner satisfaction [21], they probably had little influence on this outcome given our study design as these elements were almost identical for both groups.

The term e-learning is generic and is used to describe any kind of educational material made available through technological means. It therefore covers many different formats (video, podcast, interactive module, serious game, etc), which have been increasingly used in the last decades because of their many advantages such as flexibility, mobility, and avoidance of both time and space constraints [22]. Advanced models of e-learning can mimic real-life situations and are often considered preferable to classical models of education thanks to their interactivity. Systematic reviews have however failed to demonstrate a superiority of e-learning compared to traditional learning in improving patient outcomes or health professionals’ behaviors, skills, or knowledge [23,24]. This might be because e-learning materials are so different in conception and philosophy that they cannot be directly comparable.

Even if it cannot entirely substitute for face-to-face learning, e-learning seems to be particularly well adapted to the field of emergency medicine where shift work is common [25]. It is indeed often difficult to provide recurrent practical hands-on sessions for paramedics once they have completed their training, though they must still acquire new knowledge and skills on a regular basis. Studies have shown that e-learning is the preferred learning method for paramedics when practical training is not possible due to diverse constraints [26,27].

Stroke is the second leading cause of death worldwide and a major cause of long-term disability [28]. One of the best ways to limit stroke sequelae is to reduce the time between symptom onset and administration of the appropriate treatment. Swift and accurate identification of stroke symptoms is therefore of paramount importance to enable prompt transport to a stroke center. Some studies suggest that the use of e-learning material enhanced both neurological assessment by nurses and quality of care in stroke teams [29,30]. Different e-learning tools aiming to improve symptoms recognition and global care of stroke patients are currently being developed and implemented [31,32].

Strengths and Limitations

Though this study has several strengths, it also has limitations. First, given our initial power calculation and the small number of paramedics working in Geneva, the sample size is limited. Moreover, as there can be considerable differences between prehospital emergency medical systems or paramedic certifications, even across different Swiss regions, the results of this study might not be entirely applicable to other systems [16]. As this study only assessed learning material, it also might not adequately predict the performance of the paramedics in the field. Finally, as the video had a fixed duration of about 59 minutes, the main investigator might have been only incompletely blinded as to group allocation. Nevertheless, randomization, concealment, and blinding were performed according to the strictest standards. Comparison of this innovative highly interactive e-learning module with a well-established teaching standard is another strength of this study.

Implications for Future Research

After an update and improvement of the e-learning module used in this study, a new assessment of this teaching method should be performed. The performance of healthcare providers trained with this new method and its impact on the global care of stroke patients should also now be assessed in the field.

Conclusion

Use of an e-learning module shows promising results in teaching the NIHSS to paramedics. The module should be improved according to the study results and user feedback, and performance of healthcare providers who have followed this module should be tested in the field.

Acknowledgments

The authors would like to thank Professor Patrick Lyden for allowing the use of his original NIHSS material in this study.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was not prospectively registered. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low. The authors' argument for non-registration is that registration was not required as the study does not come into the scope of the Swiss federal law on human research. The protocol was, however, submitted to the local ethics committee.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 8873 KB - jmir_v22i6e18358_app1.pdf]

References

1. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, American Heart Association Stroke Council. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke 2018 Mar;49(3):e46-e110. [doi: 10.1161/STR.0000000000000158] [Medline: 29367334]
2. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. Arch Neurol 1989 Jun;46(6):660-662. [doi: 10.1001/archneur.1989.00520420080026] [Medline: 2730378]
3. Lyden P. Using the National Institutes of Health Stroke Scale: A Cautionary Tale. Stroke 2017 Feb;48(2):513-519. [doi: 10.1161/STROKEAHA.116.015434] [Medline: 28077454]
4. Chiu S, Cheng K, Sun T, Chang K, Tan T, Lin T, et al. The effectiveness of interactive computer assisted instruction compared to videotaped instruction for teaching nurses to assess neurological function of stroke patients: a randomized controlled trial. Int J Nurs Stud 2009 Dec;46(12):1548-1556. [doi: 10.1016/j.ijnurstu.2009.05.008] [Medline: 19552905]
5. Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati Prehospital Stroke Scale: reproducibility and validity. Ann Emerg Med 1999 Apr;33(4):373-378. [doi: 10.1016/s0196-0644(99)70299-4] [Medline: 10092713]

6. Llanes JN, Kidwell CS, Starkman S, Leary MC, Eckstein M, Saver JL. The Los Angeles Motor Scale (LAMS): a new measure to characterize stroke severity in the field. Prehosp Emerg Care 2004;8(1):46-50. [doi: 10.1080/312703002806] [Medline: 14691787]
7. Iguchi Y, Kimura K, Watanabe M, Shibazaki K, Aoki J. Utility of the Kurashiki Prehospital Stroke Scale for hyperacute stroke. Cerebrovasc Dis 2011;31(1):51-56. [doi: 10.1159/000320854] [Medline: 20980754]
8. Kesinger MR, Sequeira DJ, Buffalini S, Guyette FX. Comparing National Institutes of Health Stroke Scale among a stroke team and helicopter emergency medical service providers. Stroke 2015 Feb;46(2):575-578. [doi: 10.1161/STROKEAHA.114.007850] [Medline: 25538202]
9. Venkat A, Cappelen-Smith C, Askar S, Thomas PR, Bhaskar S, Tam A, et al. Factors Associated with Stroke Misdiagnosis in the Emergency Department: A Retrospective Case-Control Study. Neuroepidemiology 2018;51(3-4):123-127. [doi: 10.1159/000491635] [Medline: 30092562]
10. Sommer P, Seyfang L, Posekany A, Ferrari J, Lang W, Fertl E, et al. Prehospital and intra-hospital time delays in posterior circulation stroke: results from the Austrian Stroke Unit Registry. J Neurol 2017 Jan;264(1):131-138 [FREE Full text] [doi: 10.1007/s00415-016-8330-x] [Medline: 27822599]
11. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, et al. Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. Stroke 1994 Nov;25(11):2220-2226. [doi: 10.1161/01.str.25.11.2220] [Medline: 7974549]
12. Tarpada SP, Morris MT, Burton DA. E-learning in orthopedic surgery training: A systematic review. J Orthop 2016 Dec;13(4):425-430 [FREE Full text] [doi: 10.1016/j.jor.2016.09.004] [Medline: 27688638]
13. David I, Poissant L, Rochette A. Clinicians' expectations of Web 2.0 as a mechanism for knowledge transfer of stroke best practices. J Med Internet Res 2012 Sep 13;14(5):e121 [FREE Full text] [doi: 10.2196/jmir.2016] [Medline: 23195753]
14. Kang S, Kim S, Kwon Y, Kim T, Park H, Park H, et al. The virtual asthma guideline e-learning program: learning effectiveness and user satisfaction. Korean J Intern Med 2018 May;33(3):604-611 [FREE Full text] [doi: 10.3904/kjim.2017.017] [Medline: 28724281]
15. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
16. Chan M, Fehlmann CA, Pasquier M, Suppan L, Savoldelli GL. Endotracheal Intubation Success Rate in an Urban, Supervised, Resident-Staffed Emergency Mobile System: An 11-Year Retrospective Cohort Study. J Clin Med 2020 Jan 16;9(1):238 [FREE Full text] [doi: 10.3390/jcm9010238] [Medline: 31963162]
17. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 2014 Jun 10;111(23):8410-8415 [FREE Full text] [doi: 10.1073/pnas.1319030111] [Medline: 24821756]
18. Suppan L, Suppan M, Koka A, Stuby L. Mendeley Data. 2020 Feb 13. NIHSS - Paramedics dataset URL: https://data.mendeley.com/datasets/j2t5ph46nk/1 [accessed 2020-05-22]
19. Meyer BC, Lyden PD. The modified National Institutes of Health Stroke Scale: its time has come. Int J Stroke 2009 Aug;4(4):267-273 [FREE Full text] [doi: 10.1111/j.1747-4949.2009.00294.x] [Medline: 19689755]
20. Croxton RA. The Role of Interactivity in Student Satisfaction and Persistence in Online Learning. Journal of Online Learning and Teaching 2014 Jun;10(2):314-325 [FREE Full text]
21. Rouleau G, Gagnon M, Côté J, Payne-Gagnon J, Hudson E, Dubois C, et al. Effects of E-Learning in a Continuing Education Context on Nursing Care: Systematic Review of Systematic Qualitative, Quantitative, and Mixed-Studies Reviews. J Med Internet Res 2019 Oct 02;21(10):e15118 [FREE Full text] [doi: 10.2196/15118] [Medline: 31579016]
22. McGowan BS. The Rise and Stall of eLearning: Best Practices for Technology-Supported Education. J Contin Educ Nurs 2015 Jul;46(7):292-294. [doi: 10.3928/00220124-20150619-11] [Medline: 26154667]
23. Vaona A, Banzi R, Kwag KH, Rigon G, Cereda D, Pecoraro V, et al. E-learning for health professionals. Cochrane Database Syst Rev 2018 Jan 21;1:CD011736 [FREE Full text] [doi: 10.1002/14651858.CD011736.pub2] [Medline: 29355907]
24. Feng J, Chang Y, Chang H, Erdley WS, Lin C, Chang Y. Systematic review of effectiveness of situated e-learning on medical and nursing education. Worldviews Evid Based Nurs 2013 Aug;10(3):174-183. [doi: 10.1111/wvn.12005] [Medline: 23510119]
25. Roe D, Carley S, Sherratt C. Potential and limitations of e-learning in emergency medicine. Emerg Med J 2010 Feb;27(2):100-104. [doi: 10.1136/emj.2008.064915] [Medline: 20156859]
26. Knox S, Cullen W, Dunne C. Continuous Professional Competence (CPC) for Irish paramedics and advanced paramedics: a national study. BMC Med Educ 2014 Mar 02;14:41 [FREE Full text] [doi: 10.1186/1472-6920-14-41] [Medline: 24580830]
27. Sherratt FC, Snape D, Goodacre S, Jackson M, Pearson M, Marson AG, et al. Paramedics' views on their seizure management learning needs: a qualitative study in England. BMJ Open 2017 Jan 09;7(1):e014024 [FREE Full text] [doi: 10.1136/bmjopen-2016-014024] [Medline: 28069626]
28. Katan M, Luft A. Global Burden of Stroke. Semin Neurol 2018 Apr;38(2):208-211. [doi: 10.1055/s-0038-1649503] [Medline: 29791947]

29. Shin JY, Issenberg SB, Roh YS. The effects of neurologic assessment E-learning in nurses. Nurse Educ Today 2017 Oct;57:60-64. [doi: 10.1016/j.nedt.2017.07.007] [Medline: 28734188]
30. Jones SP, Miller C, Gibson JME, Cook J, Price C, Watkins CL. The impact of education and training interventions for nurses and other health care staff involved in the delivery of stroke care: An integrative review. Nurse Educ Today 2018 Feb;61:249-257. [doi: 10.1016/j.nedt.2017.11.024] [Medline: 29272824]
31. Václavík D, Bar M, Klečka L, Holeš D, Čábal M, Mikulík R. Prehospital stroke scale (FAST PLUS Test) predicts patients with intracranial large vessel occlusion. Brain Behav 2018 Sep;8(9):e01087 [FREE Full text] [doi: 10.1002/brb3.1087] [Medline: 30085409]
32. Birns J, Woodward P, Brenton H, Bello F. Development of a Novel Multimedia E-Learning Tool for Teaching the Symptoms and Signs of Stroke. Creative Education 2018 Jun;09(08):1196-1211. [doi: 10.4236/ce.2018.98089]

Abbreviations

NIHSS: National Institutes of Health Stroke Scale
SMUR: Service Mobile d’Urgences et de Réanimation

Edited by G Eysenbach; submitted 21.02.20; peer-reviewed by I David, S Sarbadhikari; comments to author 30.03.20; revised version received 01.04.20; accepted 16.04.20; published 09.06.20.

Please cite as:
Koka A, Suppan L, Cottet P, Carrera E, Stuby L, Suppan M
Teaching the National Institutes of Health Stroke Scale to Paramedics (E-Learning vs Video): Randomized Controlled Trial
J Med Internet Res 2020;22(6):e18358
URL: http://www.jmir.org/2020/6/e18358/
doi:10.2196/18358
PMID:32299792

Original Paper

The Model of Gamification Principles for Digital Health Interventions: Evaluation of Validity and Potential Utility

Mark Floryan[1], PhD; Philip I Chow[2], PhD; Stephen M Schueller[3], PhD; Lee M Ritterband[2], PhD

[1]Department of Computer Science, University of Virginia, Charlottesville, VA, United States

[2]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

[3]University of California, Irvine, CA, United States

Corresponding Author:
Mark Floryan, PhD
Department of Computer Science
University of Virginia
85 Engineer's Way
Charlottesville, VA, 22904
United States
Phone: 1 4342433087
Email: mrf8t@virginia.edu

Abstract

Background: Although gamification continues to be a popular approach to increase engagement, motivation, and adherence to behavioral interventions, empirical studies have rarely focused on this topic. There is a need to empirically evaluate gamification models to increase the understanding of how to integrate gamification into interventions.

Objective: The model of gamification principles for digital health interventions proposes a set of five independent yet interrelated gamification principles. This study aimed to examine the validity and reliability of this model to inform its use in Web- and mobile-based apps.

Methods: A total of 17 digital health interventions were selected from a curated website of mobile- and Web-based apps (PsyberGuide), which makes independent and unbiased ratings on various metrics. A total of 133 independent raters trained in gamification evaluation techniques were instructed to evaluate the apps and rate the degree to which gamification principles are present. Multiple ratings ($n \geq 20$) were collected for each of the five gamification principles within each app. Existing measures, including the PsyberGuide credibility score, mobile app rating scale (MARS), and the app store rating of each app were collected, and their relationship with the gamification principle scores was investigated.

Results: Apps varied widely in the degree of gamification implemented (ie, the mean gamification rating ranged from $0.17 \leq m \leq 4.65$ out of 5). Inter-rater reliability of gamification scores for each app was acceptable ($\kappa \geq 0.5$). There was no significant correlation between any of the five gamification principles and the PsyberGuide credibility score ($P \geq .49$ in all cases). Three gamification principles (supporting player archetypes, feedback, and visibility) were significantly correlated with the MARS score, whereas three principles (meaningful purpose, meaningful choice, and supporting player archetypes) were significantly correlated with the app store rating. One gamification principle was statistically significant with both the MARS and the app store rating (supporting player archetypes).

Conclusions: Overall, the results support the validity and potential utility of the model of gamification principles for digital health interventions. As expected, there was some overlap between several gamification principles and existing app measures (eg, MARS). However, the results indicate that the gamification principles are not redundant with existing measures and highlight the potential utility of a 5-factor gamification model structure in digital behavioral health interventions. These gamification principles may be used to improve user experience and enhance engagement with digital health programs.

(J Med Internet Res 2020;22(6):e16506) doi:10.2196/16506

KEYWORDS

gamification; internet interventions; eHealth; mHealth; digital health

Introduction

There is substantial interest in understanding how gamification can improve electronic health (eHealth) and mobile health (mHealth) interventions [1-3], yet significant gaps in the literature remain. Metareviews of gamification strategies within behavioral interventions summarize the work in this area [4,5] but often focus on the mechanics employed (eg, badges, leaderboards, etc) and neglect other potentially important factors (ie, the context of the app). In addition, individual studies of gamification of digital health interventions typically present results as general game vs control studies [6,7] or focus more on the qualitative and subjective aspects of gamification [8]. As gamification techniques are intertwined with other intervention components, their validity and incremental impact are relatively unknown.

The model of gamification principles for internet interventions [9] was developed to present a unifying, theory-driven set of five gamification principles (Textbox 1) that can be used in the building and testing of digital health interventions. There is no widespread agreement on how the principles of gamification should be applied to eHealth/mHealth interventions. Although the detailed justification for the development of these principles is beyond the scope of this paper, additional information can be found in the prior study on which this one is based [9]. These principles, which were extracted from several well-known gamification models, represent independent and actionable items regarding the application of gamification (see the study by Floryan et al [9] for details). The model is composed of five separate yet interrelated constructs: meaningful purpose, meaningful choice, supporting player archetypes, feedback, and visibility. Textbox 1 summarizes the five gamification principles and provides a short description of each principle. The principles provide concrete and measurable descriptions of gamification while focusing on the context, goals, and attitudes of users of the program. These principles provide a descriptive framework for measuring both the presence and quality of gamification implementation. Although these have been well defined, empirical validation is a necessary next step.

Textbox 1. Summary of principles of gamification.

Gamification principle and description:

- Meaningful purpose: The app presents goals that align with the user’s motivations and interests
- Meaningful choice: The app gives users agency over how they achieve their goals
- Supporting player archetypes: Mechanics in the app leverage individual user and player characteristics
- Feedback: The app communicates how user actions affect progress
- Visibility: The app makes clear to users the amount of progress made and how much more is needed

The proposed gamification principles encourage developers to separate the idea of typically considered gamification mechanics (eg, points, badges, and leaderboards) from the purpose of those mechanics (eg, motivating the purpose, increasing user choice, supporting player archetypes, etc). In this way, the model encourages researchers to consider the mechanics of gamification and how those relate to the underlying motivational affordances of the user. It also provides researchers with a framework for implementing these mechanics within the technology-based interventions they create. In behavioral science, several attempts have been made to specify various behavior change strategies [10] and call for increased use of evidence-based behavior change techniques within research and commercially developed products and interventions [11]. However, there have been few efforts to map behavior change principles to design features that could be implemented by researchers. The principles of gamification attempt to address this mapping by providing actionable principles contextualized by known behavior change principles for internet interventions (details regarding this mapping can be found in the study by Floryan et al [9]). The specification of these principles can facilitate research and lead to a better understanding of these mechanisms and how they can best be used. The goals of this study were to understand the validity and reliability of this 5-factor gamification model.

Methods

Overview

Mobile- and Web-based interventions were selected from PsyberGuide, a nonprofit endeavor that aims to provide consumers with information to aid in selecting different types of mental health apps. PsyberGuide conducts independent and unbiased reviews of mental health apps and evaluates products on three dimensions: credibility, user experience, and privacy and data security. PsyberGuide has evaluated over 200 mental health apps and is viewed as a useful standard for determining the quality of apps on these various metrics [12-14]. Given the variety of apps reviewed on PsyberGuide and the variance in scores of credibility and user experience of these apps, it provided a useful point of comparison to evaluate the gamification assessment. Specifically, the credibility and user experience scores (see the *Measures* section for more detail) for each app were useful to compare the presence of gamification with these other metrics. Apps were selected (Table 1) in February 2018 and evaluated for the presence and implementation of the five internet intervention gamification principles (Textbox 1). They were selected using the following criteria: (1) presence on the PsyberGuide listing, (2) free to use or provided core content free of charge, (3) broadly applicable to the general population in the domain of behavioral and mental health, and (4) available through at least one of the Apple App Store, Android Store, or a Web browser (eg, website based). A total of 133 trained independent raters and undergraduates enrolled in a college-level human-computer interaction (HCI) course served as judges. The raters were all aged between 18 and 22 years and majoring in computer science or a related field (eg, systems engineering, computer engineering). Many of the raters were double majoring in a related field (eg, cognitive science, psychology, etc). To obtain adequate interrater reliability, Saito et al [15] recommend a higher number of ratings when the potential variance between ratings is high, 20 to 25 ratings were desired for each app. Therefore, with each judge providing up to 3 ratings, 17 apps were selected.

Table 1. List of apps and their modality or purpose.

App name	Modality or purpose
Headspace	Mindfulness and meditation
Lumosity	Cognitive training
iSleep Easy	Meditation and restful sleep
FitBrains	Cognitive training
Happify: For Stress & Worry	Well-being and happiness
Serenita	Stress and anxiety
SuperBetter	Goal setting, resilience, motivation
Flowy	Breathing and relaxation
The Mxindfulness App	Mindfulness and meditation
HAPPYneuron	Cognitive training
Smiling Mind	Mindfulness and meditation
Pacifica (now Sanvello)	Anxiety and depression
Virtual Hope Box	Anxiety and depression
Peak	Cognitive training
Personal Zen	Stress and anxiety
BrainHQ	Cognitive training
Wildflowers	Mindfulness and meditation

Measures

Gamification Principles

A novel self-report measure was developed based on the gamification principles for internet interventions model. The measure is composed of five items, with each item assessing the principle of gamification. Items include a description of the gamification principle (eg, meaningful purpose) in question and instruct the rater to judge the presence of that principle within the intervention. Embedded within the descriptions are probing questions to help determine the extent the principle is present (eg, “Does the application allow the user to make decisions about how they reach their goal?”). One item was used to assess each gamification principle to increase the efficiency of raters and limit the response burden. Single-item scales have been used and validated to assess complex constructs such as self-esteem, job satisfaction, and personality traits [16-18]. The items were assessed on a 6-point Likert scale (0=complete absence of the gamification principle and 1-5=weak to strong presence of the principle in question; Multimedia Appendix 1).

App Quality (Mobile App Rating Scale)

The Mobile App Rating Scale (MARS; Stoyanov et al [19]) is a widely used measure of app quality that focuses on aspects of engagement, functionality, aesthetics, and information quality. MARS scores for each program included in this study were obtained from the PsyberGuide website. MARS scores are averaged from multiple independent raters who either have expertise in health interventions or psychology, technology development or design, or lived experience with intended clinical issues. Each MARS score was calculated using a combination of at least three raters. MARS scores are represented as *user experience* ratings (with a maximum score of 5.0) on the PsyberGuide website.

App Credibility (PsyberGuide Credibility)

PsyberGuide credibility ratings are meant to determine the likelihood that a given product will produce the proposed benefits. It is based on an assessment of the strength of research evidence, source of research evidence, specificity of the app, expertise of the development team, number of app store ratings, and recency of updates. PsyberGuide credibility rating scores are made by a team of trained reviewers consisting of undergraduate or masters-level students using an approval and consensus process (maximum score of 5.0). PsyberGuide credibility rating scores were obtained from the PsyberGuide website.

App Store Rating

The app store rating (Apple or Android) for each intervention was obtained. App store ratings are based on a system of stars

(0-5), with a higher number of stars indicating a greater degree of liking the app.

Procedure

Each of the 133 raters was randomly assigned to evaluate three apps. Raters were trained using a 2-part approach. The first was a 75-min training session that reviewed the theory of heuristic evaluations, a core concept in HCI. Heuristic evaluations occur when trained raters use a system and rate how well the design conforms to a set of described heuristics [20]. The heuristics used for this study were the five principles of gamification [9]. The training also involved teaching the raters about the principles of gamification, as these were the heuristics to be used for comparison when rating. Although this is not the prototypical use of a heuristic evaluation (a company, eg, would typically rate a user interface against a set of design principles), the training focused on how the process of rating gamification in apps was analogous to a classical heuristic evaluation. The apps were rated against a different set of heuristics (gamification principles), and these principles were enumerated and discussed in detail during the training. The training also provided a broad overview of digital health interventions (definition and brief examples) but did not include any detailed training in this area. To increase our confidence that the ratings were done thoughtfully and consistent with the training guidelines, raters were required to provide a written justification for their ratings.

Raters were given 2 weeks to evaluate their apps and were instructed to use each app for at least 15 min every day. Specifically, raters were asked to use the app as a normal user and to examine the presence of each of the gamification principles. Although rater usage was not tracked, the raters were encouraged to maintain lists of specific examples of each gamification principle they encountered and to list them in their written justifications. After 2 weeks, raters were given 1 week to complete a 10-question survey. For each gamification principle, raters were asked whether the principle was present in the app (binary), and, if so, to what degree that principle was present (1-5 scale). These pairs of questions were later combined to create single 6-point (0-5 scale) responses. For the presence questions (1-5 scale), a description was provided for scores 1, 3, and 5. This was done to allow raters some flexibility in interpreting the score between these endpoints and middle points. To ensure that raters had provided thoughtful responses, they were asked to provide a justification for each of their scores by writing a supporting paragraph. The raters were assigned a grade to complete this assignment and to provide reasonable and thoughtful justifications for the provided scores. Raters provided reasonable justifications, earning an average of 9.4 out of 10 on this assignment. The university institutional review board (IRB) was contacted with the details of this endeavor, and it was determined that no IRB protocol was necessary.

Statistical Analysis Outliers

Statistical analysis outliers (ratings more than two SDs from the mean) were identified. However, on review of the justifications provided by these raters, no data were removed. For each program and survey question combination, the mean scores and SDs were calculated. Interrater reliability scores for each app were calculated to determine the degree of agreement among the independent raters. Interrater reliability was obtained by using the weighted Fleiss kappa [21,22] for each app, across all questions and raters. Fleiss kappa is recommended when there are more than two raters.

Correlations and *P* values were calculated between the average ratings of each gamification principle and each of the three dependent measures (ie, app quality, app credibility, app store rating). Correlations between the gamification principles were examined to detect the presence of collinearity among the ratings. A custom program, written in Python, was used for outlier identification, coalescing the raw data into mean (SD), and for computing the interrater reliability (ie, all results from Tables 2 and 3). The statistics program R was used to compute all correlation coefficients and related statistics (ie, all results from Multimedia Appendix 2 and Table 4).

Table 2. Mean and SD ratings for each gamification principle across raters.

App name	Meaningful purpose, mean (SD)	Meaningful choice, mean (SD)	Supporting player archetypes, mean (SD)	Feedback, mean (SD)	Visibility, mean (SD)
Headspace	4.29 (0.84)	3.92 (1.04)	3.50 (1.47)	3.33 (1.49)	4.08 (1.08)
Lumosity	4.36 (0.79)	3.08 (1.49)	3.40 (1.10)	4.32 (0.93)	3.76 (1.21)
iSleep Easy	3.92 (1.00)	2.71 (1.57)	0.17 (0.47)	1.17 (1.40)	0.25 (0.60)
FitBrains	4.10 (1.38)	2.76 (1.69)	3.33 (1.04)	4.38 (1.09)	4.29 (1.39)
Happify: For Stress & Worry	4.26 (1.07)	3.61 (1.31)	3.26 (1.72)	3.83 (1.27)	3.74 (1.36)
Serenita	3.91 (1.06)	2.57 (1.50)	1.43 (1.58)	3.43 (1.47)	3.70 (1.23)
SuperBetter	3.65 (1.05)	3.61 (1.34)	3.00 (1.41)	3.26 (0.94)	3.48 (1.25)
Flowy	1.95 (1.07)	1.64 (1.37)	1.91 (1.44)	2.32 (1.49)	2.91 (1.20)
The Mindfulness App	3.59 (1.34)	3.59 (1.30)	2.41 (1.44)	2.18 (1.61)	2.86 (1.42)
HAPPYneuron	4.15 (0.91)	3.25 (1.30)	2.90 (1.34)	3.35 (1.24)	3.55 (1.16)
Smiling Mind	4.09 (1.06)	3.78 (1.18)	2.57 (1.61)	3.26 (1.80)	3.96 (0.95)
Pacifica	4.04 (1.04)	4.40 (0.98)	4.56 (0.70)	3.64 (1.29)	4.16 (0.97)
Virtual Hope Box	3.30 (1.43)	3.43 (1.38)	2.43 (1.77)	0.70 (1.16)	0.17 (0.38)
Peak	4.35 (0.91)	3.13 (1.73)	2.96 (1.60)	4.26 (0.85)	4.65 (0.70)
Personal Zen	2.14 (1.12)	0.86 (1.17)	0.95 (1.13)	2.86 (1.36)	3.43 (1.37)
BrainHQ	4.08 (1.19)	2.54 (1.68)	2.58 (1.55)	4.00 (0.82)	3.75 (1.16)
Wildflowers	3.65 (1.34)	2.78 (1.44)	2.35 (1.68)	3.52 (1.35)	3.91 (1.10)

Table 3. Interrater reliability scores for each app.

App	Interrater reliability
Headspace	0.53
Lumosity	0.57
iSleep Easy	0.67
FitBrains	0.55
Happify: For Stress & Worry	0.51
Serenita	0.53
SuperBetter	0.54
Flowy	0.52
The Mindfulness App	0.51
HAPPYneuron	0.54
Smiling Mind	0.52
Pacifica	0.58
Virtual Hope Box	0.61
Peak	0.57
Personal Zen	0.58
BrainHQ	0.54
Wildflowers	0.51

Table 4. Correlation matrix between gamification principles across all rated apps.

Gamification principle	Meaningful purpose	Meaningful choice	Supporting player archetypes	Feedback	Visibility
Meaningful purpose	N/A[a]	0.71	0.50	0.48	0.30
Meaningful choice	N/A	N/A	0.69	0.11	0.11
Supporting player archetypes	N/A	N/A	N/A	0.57	0.57
Feedback	N/A	N/A	N/A	N/A	0.92
Visibility	N/A	N/A	N/A	N/A	N/A

[a] N/A: not applicable.

Results

Overview

The means and SDs for each gamification principle across the 17 apps are shown in Table 2. There was a wide degree of variation in gamification present in the apps. The average gamification score (ie, the mean of all five gamification principle ratings) ranged from 1.64 to 4.16 out of 5. Among the principles, supporting player archetypes was judged as being the least present (average 2.57 out of 5), whereas meaningful purpose was judged as being most present (average 3.75 out of 5).

Table 3 lists the interrater reliability scores for each app studied. Interrater reliability scores range from 0.51 to 0.67, indicating acceptable levels of agreement across raters for each app [23].

Criterion Validity and Associations Between Gamification Principles and Other Measures

Table 5 lists the app credibility score (ie, PsyberGuide Credibility Score), the MARS score, and the app store ratings for each app. One app (HAPPYNueron) did not have an app store rating. In general, the three scores were not strongly associated with one another ($r=-0.09$ for PsyberGuide Credibility vs MARS; $r=0.32$ for MARS vs app store rating; $r=0.04$ for PsyberGuide Credibility vs app store rating), suggesting that each of these three measures likely represent different aspects of app quality.

Table 5. Various metrics scoring each studied app (1-5 scale).

App	PsyberGuide Credibility Score	Mobile App Rating Scale	App store rating
Headspace	4.64	4.74	4.9
Lumosity	3.21	4.34	4.7
iSleep Easy	3.55	3.01	4.6
FitBrains	2.85	4.67	3.7
Happify: For Stress & Worry	3.92	3.34	4.5
Serenita	3.2	3.2	3
SuperBetter	3.55	4.39	4.7
Flowy	2.5	4.1	4.3
The Mindfulness App	2.85	3.3	4.4
HAPPYneuron	2.5	4.15	N/A[a]
Smiling Mind	2.85	4	4.6
Pacifica	2.85	4.7	4.7
Virtual Hope Box	3.92	3.59	4.4
Peak	2.85	4.52	4.7
Personal Zen	3.95	3.77	2.6
BrainHQ	4.6	4.11	4.6
Wildflowers	2.85	4.08	4.33

[a] N/A: not applicable.

Multimedia Appendix 1 lists the correlation coefficients as well as t test statistics (2-tailed) and P values among gamification principles ratings and the PsyberGuide Credibility Score, the MARS rating, and the app store rating. There were generally weak correlations between the gamification principle ratings and the PsyberGuide Credibility Score, indicating a low degree of overlap. Supporting player archetypes ($P=.001$), feedback ($P=.01$), and visibility ($P=.008$) correlated strongly with the

MARS rating, whereas meaningful purpose (*P*=.04), meaningful choice (*P*=.002), and supporting layer archetypes (*P*=.04) correlated strongly with the app store ratings. A closer examination of the significant associations between the gamification principles and the MARS and app store ratings revealed between 25% and 52% shared variance (r-squared) among these variables, indicating that they are related yet measure different constructs.

Inter-Relationships Between Gamification Principles

Table 4 contains a correlation matrix of the relationships among gamification principles. The strength of associations varied widely (*r*=0.11 to 0.92), with visibility and feedback as the most strongly associated and overlapping principles. On average, the correlations were *r*=0.50, indicating that principles were related yet separate from one another.

Discussion

Principal Findings

This study provides empirical support for the model of the five gamification principles for internet interventions. We believe this model will help researchers develop new interventions and evaluate existing interventions that better engage users through the proper implementation and integration of gamification techniques. By evaluating the gamification principles in 17 health apps, the findings from this study indicate that the gamification principles are not redundant with existing app measures.

A weak relationship was found between the gamification principle ratings and the PsyberGuide credibility score. The PsyberGuide credibility score is based on several factors, some of which have no intuitive relationship to the gamification principles evaluated in this study. For example, one aspect of the PsyberGuide credibility score involves the amount of research funding the app had garnered, which has no direct connection with gamification. Other aspects of the PsyberGuide credibility score focus on the degree to which research is available on the efficacy of that app or the frequency of the updates to the app. Although some of these aspects, such as the frequency of updates, have been found to be useful predictors of some evaluations of apps such as expert-rated quality or user ratings [24], they would not be expected to categorize the features into the app. In sum, the lack of relationship between the credibility score and the gamification principle ratings suggests that the credibility of an app (which includes efficacy as well as other issues such as software support, input from experts, etc) is largely independent of its level of gamification.

There were significant relationships between 3 of the 5 gamification principles (supporting player archetypes, feedback, and visibility) and the MARS score. Theoretically, one would imagine some overlap between our gamification model and user experience aspects such as engagement. The MARS (collected from PsyberGuide) [19] explicitly mentions qualities that overlap with feedback and visibility, namely, items such as *quality or quantity of information* or *visual information*. However, the gamification principles go beyond the MARS by providing specific guidelines for presenting this information and contextualizing it within a user's broader goals and understanding. Thus, even with some conceptual overlap, the gamification principles still have added value. Engagement and attrition have long been identified as an issue within eHealth [25-27], and this can be helped by having patients play a more active role in their own care [28]. Gamification principles may therefore facilitate the execution of game mechanics by providing researchers another avenue to explore and measure engaging features that involve the patient. Similarly, the gamification principle of *supporting player archetypes* has an intuitive overlap with MARS items involving engagement and subjective quality. However, the gamification principle of supporting player archetypes presents specific mechanisms through which these qualities can be achieved and are commonly done in games and game-like systems. Thus, we believe that our gamification principles are not in direct conflict with the MARS; rather, they provide a roadmap for ways to increase app quality.

Three different gamification principles (meaningful purpose, meaningful choice, and supporting layer archetypes) were significantly associated with the app store ratings. In contrast to the MARS, the app store ratings are single overall ratings provided by end users. By directly sampling from end users, the app store rating may be viewed as largely a reflection of user choice. There was a strong association between app store ratings and the gamification principle of meaningful choice (*r*=0.71), suggesting that users may value having agency in how they use and navigate through an app. App store ratings are subjective, personal, and nonstandardized and have been shown to be an indication of app popularity, but not clinical outcomes [29]. Thus, it is notable that the strongest related principles of meaningful purpose (the user has a goal in using the app), meaningful choice (the user has agency over their progress), and supporting player archetypes (the app leverages individual user characteristics) all directly involve the user, whereas the other two gamification principles, feedback and visibility, relate more specifically to the app design, its presentation, and organization of information.

This study extends the existing literature that aims to understand technology-based behavioral interventions by identifying and coding their features [24,30,31]. Although past work has used different conceptual models, this study evaluates features based on gamification principles. Our findings that gamification principles overlap with some, but not all, other assessments of app quality and popularity indicate both convergent and discriminant validity of the gamification principles assessment. Future work may wish to apply this assessment to other behavioral change interventions and examine if gamification principles in apps correlate with real-world engagement or effectiveness. In addition, researchers and developers would be well served with a streamlined evaluation method for incorporating gamification (or measuring the presence of gamification) in their apps. Future work should focus on providing such artifacts and continuing to study their utility.

Strengths and Limitations

Several potential limitations exist for this study. Most notably, although the raters were trained with several modes of material, they were undergraduate students and not experts; however, several results from the study help limit this concern. The interrater reliability scores (weighted kappa) were all within the acceptable range (κ>0.50), which suggests that although there was variance in the scores, the raters generally gave similar ratings to apps. In addition, raters provided written justifications for each of their scores. An expert read through these justifications with the intention of removing ratings that included clear evidence of a poor rating. In the end, no ratings were deemed to lack sufficient justification, and all ratings were included in the analysis presented here.

Owing to the difficulty in calculating internal consistencies and establishing construct validity for single-item measures (to assess each gamification principle), future work should examine ways to assess gamification using more items. Although our decision was influenced by a desire to reduce the burden on raters and increase the efficiency with which ratings could be completed, there are many ways to assess gamification that should be explored in future work. However, because the results show evidence of both convergent and discriminant validity, there is some evidence that construct validity does exist.

Potential bias also exists with the app selection methodology. Although the PsyberGuide listing likely contained more than 17 apps that fit the inclusion criteria, this was deemed a sufficient number, given the quantity of raters and associated ratings to be obtained. An attempt was made to include a heterogeneous sample of apps that covered a variety of areas of focus and population types. However, a more systematic approach could have been used to select the apps, or, with sufficient resources, to include all apps.

There were some strong associations among gamification principles. Most notably, the principles of feedback and visibility had a strong relationship (r=0.92), potentially suggesting that these principles may measure the same underlying construct. It is possible that although these principle definitions are indeed mutually exclusive (ie, feedback involves the effects of user actions on the future, whereas visibility shows the results of accomplishments from the past), perhaps apps tend to use them in unison as they complement one another. It is also possible that the raters did not fully understand the distinction between these principles and may have conflated them.

Strong correlations occurred among other gamification principle ratings as well, most of which are less easily explainable. For example, the principles of meaningful choice and meaningful purpose correlate strongly (r=0.71), suggesting raters may have interpreted the meaningful qualifier as being shared across the ratings (ie, to make a meaningful choice, there must be a meaningful purpose toward which the user is progressing). More research is necessary to determine the nature of these correlations. In addition, no analysis was done to compare apps of similar purpose (eg, comparisons within mindfulness apps). Other studies have focused on comparing MARS scores or other measures among similar apps [32-35]. There might exist patterns that are stronger or weaker within apps of a specific purpose that could provide additional insight into the role of gamification and other measures (credibility rating, MARS, and app store rating), and this might be an avenue for future research.

Conclusions

In short, this paper is the first evaluation of a method determined to assess previously proposed gamification principles [9]. Our findings suggest that gamification principles relate to some, but not all, previously proposed methods of assessing app quality and popularity, which blend ratings made by experts (such as the PsyberGuide credibility scores), ratings made by consumers (app store ratings), and ratings made by both (in this case, the MARS scores). The pattern of relationships has considerable face validity, including those with the MARS scores and user ratings, and the lack of relationship with PsyberGuide credibility scores. This lack of rating does not indicate that either scale is invalid, but instead that the rating of gamification principles and credibility might offer unique perspectives in terms of understanding apps. The demonstration of a process of rating these products through collaborative assessments of a team of lightly trained raters also demonstrates a potential way to easily and scalably understand the growing number of technology-based behavioral interventions that are being developed. We also believe the demonstration that gamification principles have value helps support the application of these principles to the design of novel technology-based behavioral interventions and might help developers incorporate evidence-based behavior change strategies into their products. As such, both the methodological and conceptual contributions of this work can move forward the research and practice of gamification principles being thoughtfully applied to behavior change digital interventions.

Acknowledgments

The authors would like to acknowledge Jodie Ryu, who managed the curating of the apps, assignments of raters to apps, and other logistical matters.

Conflicts of Interest

SS receives funding from One Mind, of which PsyberGuide is a proprietary project. SS is the executive director of the PsyberGuide.

Multimedia Appendix 1
Gamification Principles and associated questionnaire used by the raters in this study.
[PDF File (Adobe PDF File), 46 KB - jmir_v22i6e16506_app1.pdf]

Multimedia Appendix 2

Correlation coefficients, t test statistics, and P values between each gamification principle's average ratings and each of the PsyberGuide credibility rating, Mobile App Rating Scale, and app store rating.

[XLSX File (Microsoft Excel File), 10 KB - jmir_v22i6e16506_app2.xlsx]

References

1. Aebli A. Tourists' motives for gamified technology use. Ann Tour Res 2019 Sep;78:102753. [doi: 10.1016/j.annals.2019.102753]
2. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
3. Yen BT, Mulley C, Burke M. Gamification in transport interventions: Another way to improve travel behavioural change. Cities 2019 Feb;85:140-149. [doi: 10.1016/j.cities.2018.09.002]
4. Hamari J, Koivisto J, Sarsa H. Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: HICSS'14; January 6-9, 2014; Waikoloa, HI, USA. [doi: 10.1109/hicss.2014.377]
5. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
6. Buffel C, van Aalst J, Bangels A, Toelen J, Allegaert K, Verschueren S, et al. A web-based serious game for health to reduce perioperative anxiety and pain in children (CliniPup): pilot randomized controlled trial. JMIR Serious Games 2019 Jun 1;7(2):e12431 [FREE Full text] [doi: 10.2196/12431] [Medline: 31199324]
7. Joyner D, Wengreen H, Aguilar S, Madden G. Effects of the FIT game on physical activity in sixth graders: a pilot reversal design intervention study. JMIR Serious Games 2019 Jun 18;7(2):e13051 [FREE Full text] [doi: 10.2196/13051] [Medline: 31215508]
8. Farič N, Yorke E, Varnes L, Newby K, Potts HW, Smith L, et al. Younger adolescents' perceptions of physical activity, exergaming, and virtual reality: qualitative intervention development study. JMIR Serious Games 2019 Jun 17;7(2):e11960 [FREE Full text] [doi: 10.2196/11960] [Medline: 31210135]
9. Floryan MR, Ritterband LM, Chow PI. Principles of gamification for internet interventions. Transl Behav Med 2019 Nov 25;9(6):1131-1138. [doi: 10.1093/tbm/ibz041] [Medline: 30944933]
10. Cane J, Richardson M, Johnston M, Ladha R, Michie S. From lists of behaviour change techniques (BCTs) to structured hierarchies: comparison of two methods of developing a hierarchy of BCTs. Br J Health Psychol 2015 Feb;20(1):130-150. [doi: 10.1111/bjhp.12102] [Medline: 24815766]
11. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83 [FREE Full text] [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
12. Lipczynska S. Psyberguide: a path through the app jungle. J Ment Health 2019 Feb;28(1):104. [doi: 10.1080/09638237.2017.1417574] [Medline: 29265897]
13. Perry K, Shearer E, Sylvers P, Carlile J, Felker B. mHealth 101: an introductory guide for mobile apps in clinical practice. J Technol Behav Sci 2019 Jun 13;4(2):162-169. [doi: 10.1007/s41347-019-00108-8]
14. Magee JC, Adut S, Brazill K, Warnick S. Mobile app tools for identifying and managing mental health disorders in primary care. Curr Treat Options Psychiatry 2018 Sep;5(3):345-362 [FREE Full text] [doi: 10.1007/s40501-018-0154-0] [Medline: 30397577]
15. Saito Y, Sozu T, Hamada C, Yoshimura I. Effective number of subjects and number of raters for inter-rater reliability studies. Stat Med 2006 May 15;25(9):1547-1560. [doi: 10.1002/sim.2294] [Medline: 16143966]
16. Robins RW, Hendin HM, Trzesniewski KH. Measuring global self-esteem: construct validation of a single-item measure and the Rosenberg self-esteem scale. Pers Soc Psychol Bull 2001;27(2):151-161. [doi: 10.1177/0146167201272002]
17. Wanous JP, Reichers AE, Hudy MJ. Overall job satisfaction: how good are single-item measures? J Appl Psychol 1997 Apr;82(2):247-252. [doi: 10.1037/0021-9010.82.2.247] [Medline: 9109282]
18. Spörrle M, Bekk M. Meta-analytic guidelines for evaluating single-item reliabilities of personality instruments. Assessment 2014 Jun;21(3):272-285 [FREE Full text] [doi: 10.1177/1073191113498267] [Medline: 23996850]
19. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
20. Nielsen J, Molich R. Heuristic Evaluation of User Interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI'90; April 1-5, 1990; Seattle, Washington, USA p. 249-256. [doi: 10.1145/97243.97281]
21. Banerjee M, Capozzoli M, McSweeney L, Sinha D. Beyond kappa: a review of interrater agreement measures. Can J Stat 1999 Mar;27(1):3-23. [doi: 10.2307/3315487]

22. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
24. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
25. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
26. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
27. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
28. Coulter A. Patient engagement--what works? J Ambul Care Manage 2012;35(2):80-89. [doi: 10.1097/JAC.0b013e318249e0fd] [Medline: 22415281]
29. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537. [doi: 10.1016/j.cbpra.2018.01.002]
30. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided ehealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
31. Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. Transl Behav Med 2018 Sep 8;8(5):793-798. [doi: 10.1093/tbm/ibx037] [Medline: 29471424]
32. Choi YK, Demiris G, Lin S, Iribarren SJ, Landis CA, Thompson HJ, et al. Smartphone applications to support sleep self-management: review and evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790 [FREE Full text] [doi: 10.5664/jcsm.7396] [Medline: 30353814]
33. Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
34. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
35. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]

Abbreviations

eHealth: electronic health
HCI: human-computer interaction
IRB: institutional review board
MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by G Eysenbach; submitted 04.10.19; peer-reviewed by CY Lin, M Lozano-Lozano, DW Yoo; comments to author 27.11.19; revised version received 22.01.20; accepted 27.02.20; published 10.06.20.

Please cite as:
Floryan M, Chow PI, Schueller SM, Ritterband LM
The Model of Gamification Principles for Digital Health Interventions: Evaluation of Validity and Potential Utility
J Med Internet Res 2020;22(6):e16506
URL: https://www.jmir.org/2020/6/e16506
doi:10.2196/16506
PMID:32519965

Original Paper

Using Geocaching to Promote Active Aging: Qualitative Study

Silvia Fornasini[1,2], MSc; Marco Dianti[1], MSc; Alessandro Bacchiega[1], BD; Stefano Forti[1], MSc; Diego Conforti[3], MSc

[1]Fondazione Bruno Kessler, Trento, Italy
[2]Department of Psychology and Cognitive Science, University of Trento, Trento, Italy
[3]TrentinoSalute 4.0, Trento, Italy

Corresponding Author:
Silvia Fornasini, MSc
Fondazione Bruno Kessler
Via Sommarive 18
Trento, 38123
Italy
Phone: 39 0461 314140
Email: s.fornasini@fbk.eu

Abstract

Background: Over the past few years, the development of technologies supporting active aging has been increasing. Among the activities that promote physical exercise by using technologies is geocaching—a treasure hunt of sorts in which participants use a receiver GPS to hide or find real or virtual objects. Although this activity is particularly suited to the promotion of healthy lifestyles in older people, geocaching remains to be unexplored in this area.

Objective: This study aimed to investigate the effectiveness of activities combining geocaching and self-tracking technologies to promote active aging, evaluate the usability of technologies, and explore the ways in which technologies have been integrated in the organizational context under examination to determine the user experience of participants.

Methods: A group of individuals aged 65 years and older (N=14) attending a senior center managed by a social cooperative was involved in the study. Some of them created the geocaching contents, and others, split into 2 teams, participated in the game. Each participant was given a pedometer bracelet and the geocaching app. The steps taken by individual participants along with the number of caches found by each group translated into team scores.

Results: The main results of the study were as follows: (1) activities in favor of active aging that involve the use of new technologies can foster the participation of elderly people; in particular, adding gamification to self-tracking can be a valid strategy to promote physical exercise among the elderly; (2) for this to happen, involvement of older people firsthand is crucial, and there must be a focus on their active involvement and empowerment in every phase of the project; and (3) the mediation of conflicts and competition that arise from the gamification could only take place because of the strong support of the organization in the form of social workers.

Conclusions: The results show that promoting active aging through technologies requires more effort than simply using these tools; it requires a wider process that involves an articulated organizational network with heterogeneous actors, technologies, and relations.

(J Med Internet Res 2020;22(6):e15339) doi:10.2196/15339

KEYWORDS

geocaching; active ageing; elderly empowerment; qualitative methods

Introduction

Challenging Aging Through Technology

Population aging involves an increase in the demand for care and an overload of the pension and welfare system. For this reason, aging can be considered a challenge for society [1]. To overcome this challenge, the promotion of active aging is essential, and one key strategy is the promotion of physical exercise. Regular physical activity delays functional decline, reduces the risk of chronic diseases and falls, improves the quality of life, and helps elderly people remain independent for as long as possible [1,2]. Over the past few years, the development of technologies supporting physical activity has been increasing, and the effectiveness of these tools in the promotion of active aging has also been demonstrated. Several studies have underlined the beneficial impact of technologies on the daily lives of this segment of the population. Technologies allow for an increase in physical activity [3,4] and combat loneliness by increasing sociality, self-esteem, and cognitive functions in the elderly [5]. With regard to the type of technologies used, the literature is divided into 2 main strands: on the one hand, studies that deal with devices for self-tracking of physical exercise, such as pedometer apps and wristbands [6,7], and, on the other hand, a recent trend has revealed virtual physical training through exergames [8,9] and serious games [10,11], making gamification an effective tool to motivate older people to exercise.

From Self-Tracking to Gamification

The most recommended physical activity for the elderly is walking. Independence in a person's life is strongly based on the ability to walk, an activity that allows one to carry out most daily tasks [12]. Emerging persuasive technology and ubiquitous wearable sensors offer much promise for improving health and fitness practices. These technologies can persuade the individual user to change his/her lifestyle and integrate a greater amount of physical activity into their daily life. Persuasive technologies use varied strategies for influencing behavior and activities, most notably self-tracking. Research has made great strides in understanding several aspects of these technologies, such as how their design affects activity and behavior, how to use visualization to motivate activities and provide awareness, and how feedback is understood and used [13,14]. A variety of monitoring devices have been studied and analyzed for their persuasive influence on practices and behavior. One of the most common devices is the pedometer, which is used to measure and motivate physical activity and also to set personal step goals [15]. With regard to elderly people, the literature highlights how technologies allow for an increase in physical activity levels and an improvement in self-esteem and cognitive functions [4,16]. However, some studies have reported that only 26% of these tools are used twice and that 33% of owners abandon the use of sensors after 6 months [17]. To enhance self-tracking mechanisms, gamification strategies have therefore become widespread. These include using rewards and incentives to support users' motivation and involving users in identifying and establishing their own goals [18-20]. Some studies have reported how the use of motivational elements such as rewards, feedback and progress bars, and competition can encourage older people to use technology and slow down cognitive aging. This highlights the reasons why they adopt, continue, or abandon such practices [21-23]. The literature on how the context of sport helps the elderly negotiate the aging process shows how competition in physical activity helps in redefining aging in terms of physical competency, social engagement, and mental stimulation [24].

Other studies have highlighted the relationship between user motivations and social behavior in the social context of self-tracking technologies [25-27]. For example, a study by Spillers and Asimakopoulos [27] explores how self-tracking is not just meaningful in a rational or an instrumental utilitarian sense but also in the sense of being a source of joy and pleasure for the individual. To account for the meaning of self-tracking, Spillers and Asimakopoulos [27] proposed conceptualizing self-tracking as a social and cultural practice that is fundamentally communicative. The combination of competition, communication, and cooperation consistently led to higher levels of intrinsic motivation [28,29].

Geocaching for Active Aging

Among the activities that promote physical exercise by using technologies, geocaching, a treasure hunt of sorts in which participants, called *geocachers*, use a receiver GPS to hide or find real or virtual objects (called caches), is one of them. Several studies have highlighted the benefits of this activity for the elderly. Studies that have addressed geocaching from an active aging perspective report the positive effects of spending time outdoors on health, well-being, and sleep [30,31]. Geocaching also stimulates group physical exercise, emphasizing the dimensions of social relations and interactions. It can therefore become a means of communication with family and friends, thus reducing the effects of social isolation that are often involved in aging [31,32]. Other studies have reported how geocaching allows older people to stay informed in an ever-changing environment and share an interest in a city's historical and artistic heritage [33]. It also provides the opportunity to learn and apply new knowledge in the field of Information and Communications Technology such as the use of digital devices and geolocation [30,34-36]. Geocaching is particularly suited for the promotion of healthy lifestyles in older people, because it combines outdoor activity with new technologies. However, the literature tends to emphasize the social, relational, and recreational aspects of the game, and how geocaching could effectively motivate elderly people to stay active is still unexplored.

Objectives of the Study

The Impronte project was created in 2017 as part of a series of initiatives aimed at promoting healthy lifestyles by the TrentinoSalute 4.0 Competence Center, which includes the Department of Health and Social Solidarity of the Province of Trento, the Provincial Health Services, and the Bruno Kessler Foundation. This study, which involved the senior center *Contrada Larga* of Trento, Italy, managed by the Kaleidoscopio social cooperative, aimed to show the results of a geocaching activity combined with the use of pedometer bracelets aimed to promote active aging. There were 3 main objectives of the study: first, to investigate how geocaching could effectively motivate elderly people to stay active, in addition with combining a game with the use of self-tracking technologies to quantify physical activity; second, to evaluate the usability of technologies; and third, to explore the ways in which technologies have been integrated into the organizational context under examination to determine the user experience of participants.

Methods

Context

To involve participants aged 65 years and older, the service center for elderly people, *Contrada Larga* of Trento city, was contacted. The center is managed by the *Kaleidoscopio* social cooperative that has encouraged activities concerning active aging for years. The service center is considered to be a meeting point for self-sufficient elderly people who use the spaces to meet, find and exchange useful information, have discussions with operators and other people, read newspapers, and use computers. The project activities were therefore fitted into a context that was particularly attentive: care was taken to promote the active participation of the elderly through their involvement in the ideation and realization of the activities. Thanks to the intervention of the center operators, 14 people aged between 65 and 82 years, 13 women and 1 man, were involved (Table 1).

Table 1. Demographics of study participants.

Demographics	Number of participants
Age (years)	
60-70	6
71-80	7
Above 80	1
Gender	
Male	1
Female	13
Level of education	
Primary school	N/A[a]
Secondary school	3
Tertiary school	7
Degree	4

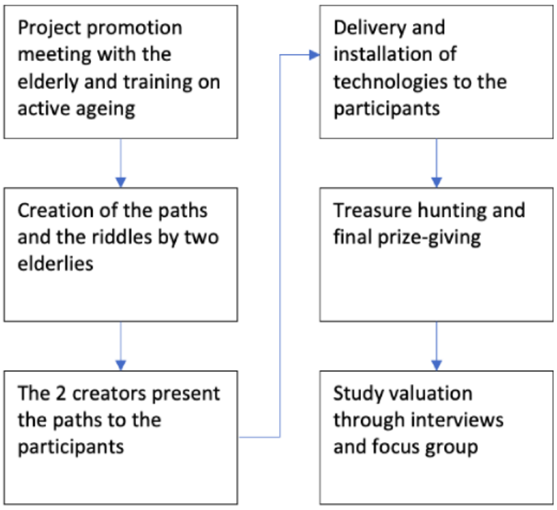
[a] N/A: not applicable.

Study Design

The project took place between April and May 2017 and was divided into 6 main phases (Figure 1):

1. First meeting: project activities, methods, and roadmap were presented to the participants. The meeting was an opportunity to offer a training introduction on active aging and on the importance of adopting a healthy lifestyle.
2. Ideation of the treasure hunt by two of the participants who collaborated in the creation of the contents of the game, identifying culturally or historically significant places in the local area.
3. Peer-to-peer presentation of the treasure hunt. A description, by researchers, of the functionalities of the technologies and the score collection mechanism.
4. The delivery of pedometer bracelets to participants, installation of the geocaching app on the participants' smartphones, team formation, and designation of team captains.
5. Treasure hunt and final distribution of prizes.
6. Study evaluation through interviews and focus groups.

Figure 1. Flowchart of project phases.



The System

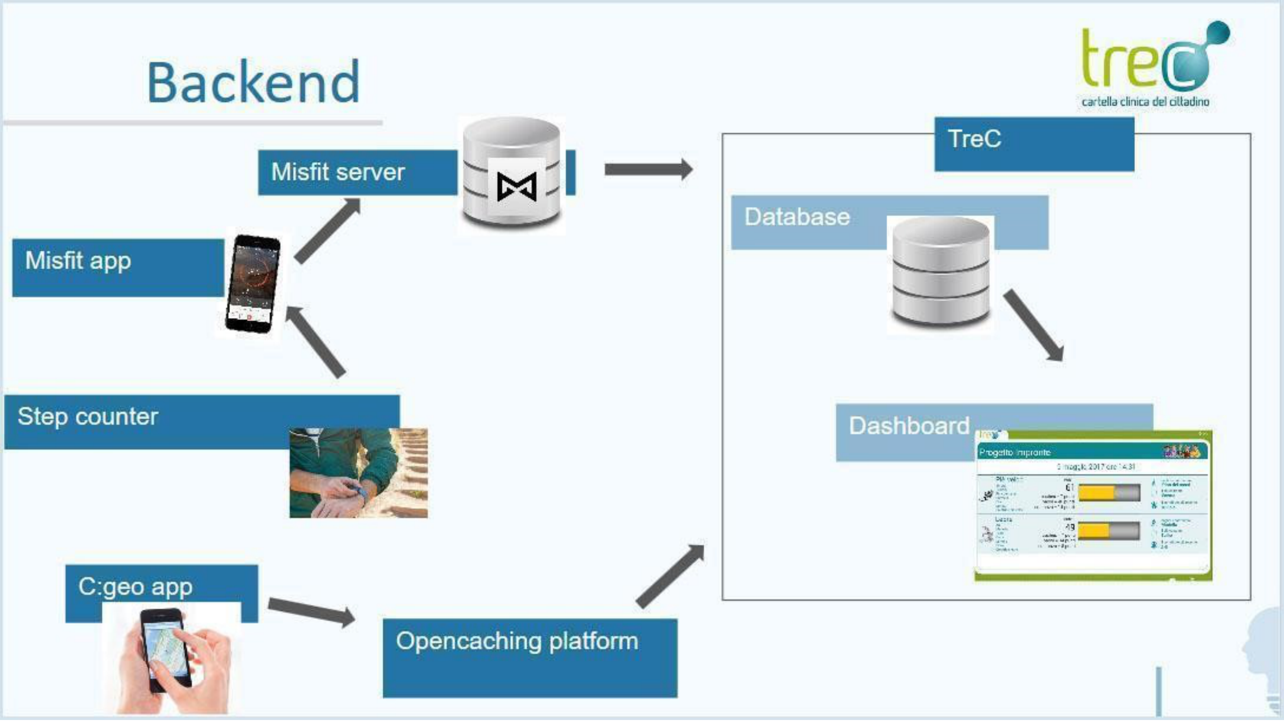
The idea behind the project was to combine the geocaching activity, accompanied by some gamification elements, with the use of fitness tracking technology in the form of a pedometer bracelet. The steps taken by individual participants along with the number of caches found by each group translated into team scores. Each participant had an activity goal of 10,000 steps a day, as advised by the World Health Organization [2]. The project therefore entailed the creation of a system that would connect the technological devices required for geocaching with those necessary for gamification: the opencaching.de website; the C: geo app; and the pedometer bracelet, a digital dashboard for step tracking and reporting of the caches found (Figure 2).

The *opencaching.de* website was selected as we preferred an open-access platform. Unlike other geocaching platforms, it is created by users, allowing anyone to register and play. It also provides the application programming interfaces (APIs) needed to access the data (number of finds, position, comments, etc). The APIs were critical to load the score on the dashboard and keep it updated daily during the challenge. To keep the score of each team, the total number of caches found was included, which we accessed from the opencaching.de website in the caches_found attribute.

The *C: geo app*, available for Android, was selected for the cache search. The app is free and is among the most complete in terms of functionality and compatibility with geocaching services. Being fully compatible with the open caching platform, it was the tool that allowed the teams to read the riddles from the open caching platform, find the caches, log them, and leave comments.

With regard to self-tracking technologies, we chose to give to each participant a pedometer bracelet. The score of each participant as well as that of the team was visible on the digital monitoring dashboard and was automatically calculated based on the number of steps taken, the number of caches found during the outings, and the number of days in which users walked more than 10,000 steps. For both teams, the scores ranged from 0 to 100, calculated according to the following proportions: cache score, 40%; step score, 40%; and constancy score, 20%. The cache score was given by the number of caches found, and the step score was 100 for the team with the highest number of total steps and a proportionally lower number for the other team. The last component of the score, *constancy*, was determined as a proportion between the number of days elapsed since the start of the race and the number of days in which the team members exceeded 10,000 steps. The data were made available periodically to the participants at the senior center or on the dedicated website.

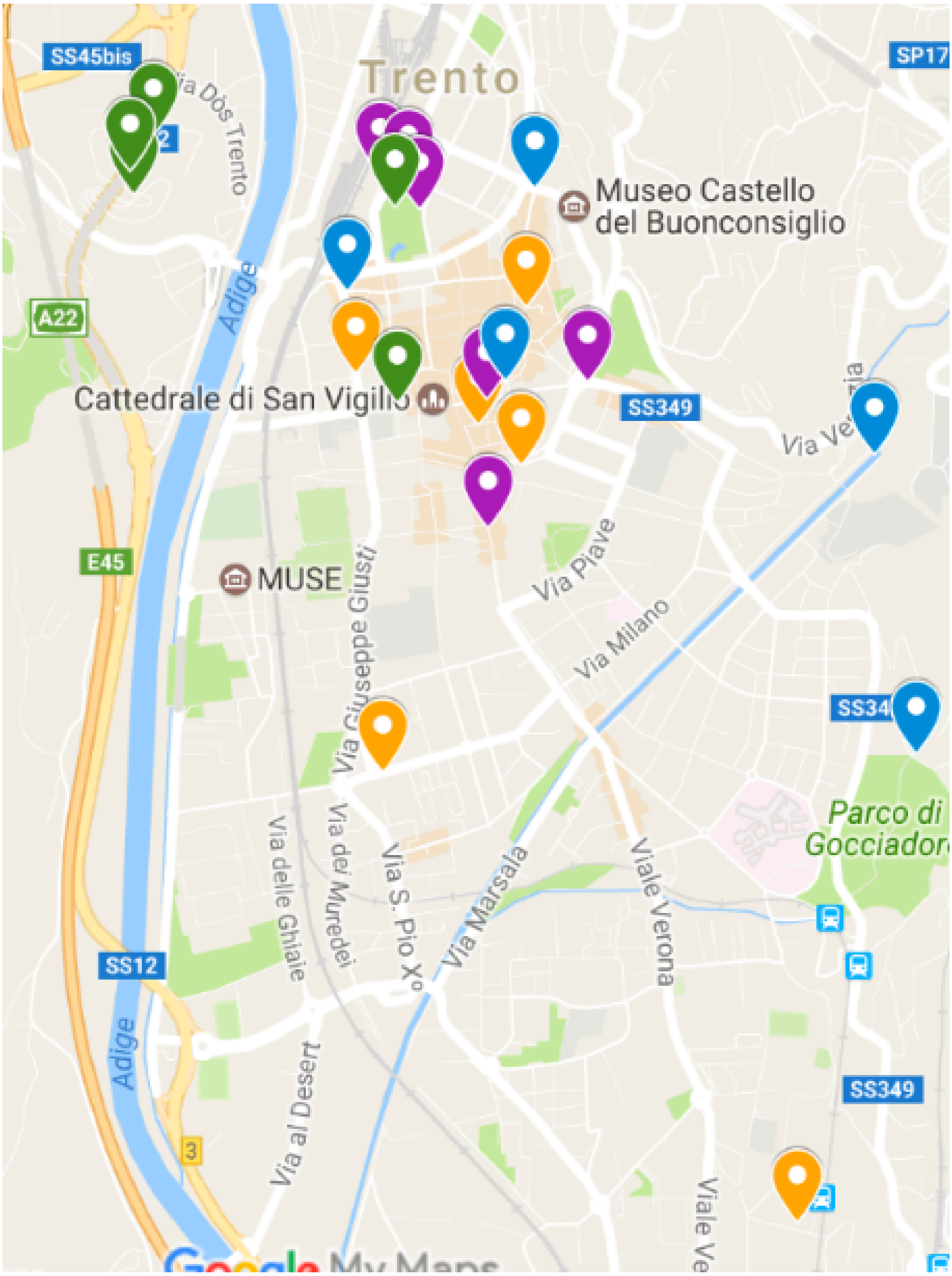
Figure 2. The backend system to collect the steps and the data from the C:Geo app and to show them in the dashboard.



Building Active Aging in a Participatory Perspective

One of the main characteristics of the project was an approach that would allow the elderly to participate actively in most of the project phases. For this reason, the treasure hunt was created by two of the participants who identified the stages of the hunt and the riddles to be solved without the help of the researchers. The 2 *creators*, after taking a tour of the city center and its surroundings, decided to design a treasure hunt divided into 4 thematic routes: "The city over the centuries," "The Italian hallmarks of the city," "Water," and "Buildings and monuments." For each route, the 2 creators identified 5 or 6 historically, artistically, and culturally relevant points (Figure 3). For each point, they created a card with its description, some curiosities, and a riddle, which could only be solved by visiting the place. In total, 24 points of interest were designed. The cards were then shared and discussed with the researchers, who uploaded them to the opencaching.de website, thus creating the actual virtual caches. The 2 creators had the task of describing the characteristics of the routes to the group of participants during a meeting that preceded the treasure hunt. During this meeting, the duration, modalities, and characteristics of the routes were presented to the group of seniors, without excessively disclosing the details to keep the curiosity alive. Under the supervision of the center operators, the participants split into 2 teams and appointed 2 team leaders with the task of using the C: geo app during the geocaching activity.

Figure 3. The points of the treasure hunter in the city of Trento.



The Treasure Hunt and Final Awarding

A total of 4 treasure hunts were created, 1 per week. The 2 teams would meet at the senior center half an hour before each hunt to prepare and organize the search, following which the teams would set out on the search to reach the greatest number of stages in the set time. We had to decide whether to encourage learning about technology by giving each participant a smartphone with the app and risk-isolated participation or whether to encourage the social aspect of the game by stimulating communication between participants. We chose the second option, and each team was accompanied by an operator. The senior center was the departure and arrival point for the participants. Both team leaders had a smartphone equipped with the C: geo app and were tasked with guiding the team's search for the stops.

Before each hunt, we recommended that participants try to leave comments and photographs in the app at each stage. This was done because one of the recreational and relational aspects of geocaching is the possibility to contribute to the contents of the game firsthand. We decided to try to focus the activity also on this aspect because the literature on elderly competition in physical activity shows that the combination of competition and cooperation consistently led to higher levels of motivation. By commenting on the cache and taking pictures, participants would have the opportunity to collaborate more. The 2 creators of the hunt, knowing the solutions to the riddles, were not able to participate in the outings, but they contributed nonetheless to the score of their team—they did not know to which team they had been assigned until the end—by accumulating steps. The final event was an opportunity to sum up the outcomes of the experience as well as to award the prize to the winning team. Each participant was given a certificate of participation.

The Evaluation

The evaluation of the study was aimed at investigating the effectiveness of the activity, the usability of the system, and the relevance of the organizational context in which the activity was organized. The first evaluation of the effectiveness of the activity was based on quantitative data: the average daily steps and the constancy (Figure 4) in doing 10,000 steps daily (x-axis: data and y-axis: number of steps). These data show a good level of engagement: the average daily steps per participant was 9660; the constancy oscillated between 4562 steps and 15,854 steps, as shown in the figure.

The second evaluation was based on qualitative data. The small number of participants involved in this first pilot study and the exploratory nature of the project led to the use of semistructured interviews and focus groups to collect the data. A total of 3 semistructured interviews were conducted with the operators and the coordinator of the center to examine in depth the characteristics of the context being studied and the perceptions of the stakeholders with regard to the project. The interviews had a duration of about an hour each. They were recorded and later transcribed. A total of 2 focus groups were used to involve the participants. The focus group is a particular type of focused meeting or interaction [37] that occurs when people recruited by a research group cooperate to turn their attention to a topic of discussion presented by a facilitator [38]. The reduced number of participants ensures and facilitates interaction [39]. Some authors [40] suggest the use of this technique to collect the opinions of older people on technology by stimulating a conversation about ideas or themes, bringing out important data that are difficult to investigate through face-to-face interviews [41]. The 2 focus groups involved 7 and 5 seniors. The first group discussion was attended by 3 members of a team and 4 members of the other team, a creator, and no team leader. The second group discussion saw the presence of 3 people from one team and 2 people from the other team, including the 2 team leaders. Both focus groups, which took place at the senior center, were led by a facilitator. The questions were asked to the whole group, and participants were encouraged to respond freely. The facilitator prepared a discussion guide with the questions (Table 2), which were considered "moments of an interaction ritual" [38] and therefore used flexibly and adapted to the conversation flow. Following an opening speech that described the plan for the meeting to the participants, they were asked questions about the organizational context of the project, the reasons that led them to participate, the usability of the technology, and the critical aspects of the gamification and finally some suggestions to improve the activity.

Figure 4. Daily steps constancy.

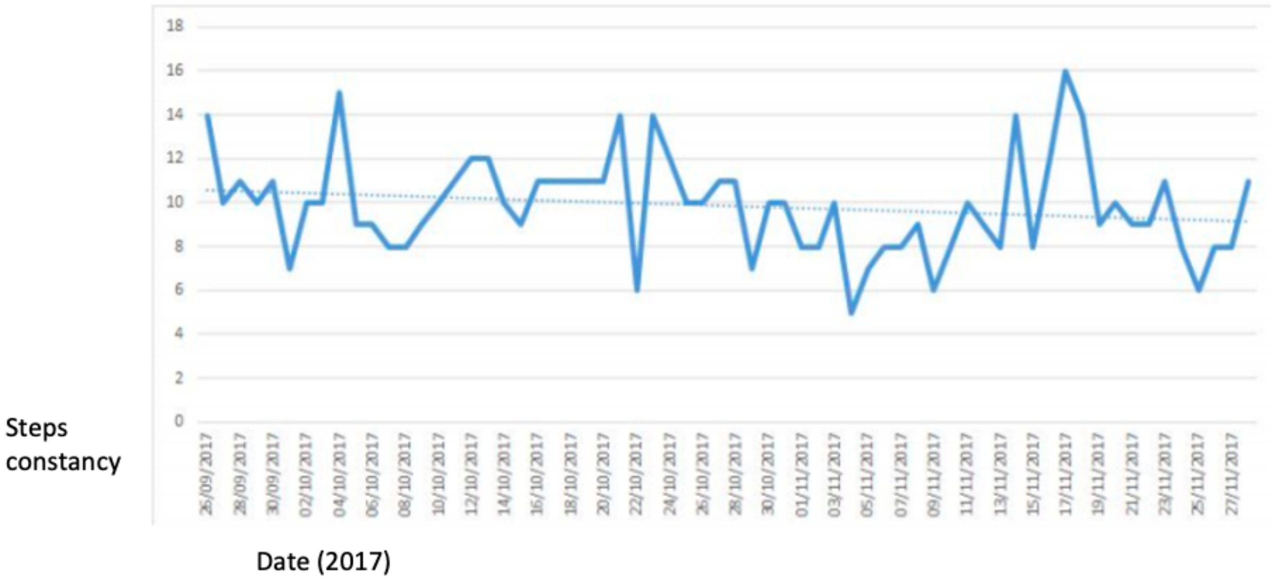


Table 2. Focus group discussion guide.

Stage	Question
Opening question	• As you know, you have been asked to participate in a research project that looks at active ageing through games and the use of new technologies: today, I would like to discuss with you how the project went.
Introductory questions	• How did you come across this project? (eg, flyer, word of mouth, operators' invitation). • What do you think about this project compared to other initiatives offered by the centre? • Do you deem it appropriate for a centre to offer a larger number of initiatives involving physical exercise and technology?
Transitional questions	• What were the reasons that led you to participate in the project? (eg, curiosity about technology, desire to take more exercise, company) • Were there people who did not participate? Why?
Key questions	• Since I was present only at some stages of the project, would you please tell me how the hunts worked? • What suggestions would you give me in order to create interesting paths? (this question was made by the participant who created the path) • Would you check your progress on the step counting app often? (eg, during the hunt, during day-to-day activities) • What are, in your opinion, the reasons why you did not take pictures or make comments? • Would you find yourselves checking on the progress of the other team on the website? And on the screenshot posted at the centre? • Would you talk about it during the week? (Was there anyone who motivated the others? Would you talk about it only at the meetings at the centre?) • What were the pros and cons of playing in a team? What suggestions would you give me in order to form teams that work well together?
Final question	• Should we propose these activities in other contexts, and do you have any suggestions to improve it?

Qualitative Data Analysis

With regard to the phase of transcription and analysis of the material collected during the focus groups, we adopted the *content analysis* or *template analysis* technique [4], a dynamic form of analysis oriented to summarizing the information content of data collected [42,43]. In particular, we opted for an *inductive* analysis of content, a "research method for the subjective interpretation of the content of text through the systematic classification process of coding and identifying themes or patterns" [44]. In the inductive analysis of the contents, categories derive directly from raw data: the use of prebuilt categories was therefore avoided [44-46]. The validity of the inference is ensured by a systematic codification process, which begins with the isolation of fragments of text by grouping them into common themes. For each topic, the different opinions of the participants were then identified, focusing on the ways in which they emerged and on the interactions with each other. In addition to the analysis of the verbal data content, the situational factors of the focus groups were considered [47], in particular, the interaction between the participants and the forms of social influence that occur in the group [38].

Results

The inductive analysis of contents and interactions revealed 3 main categories: *gamifying self-tracking: motivation vs competition*, *gamifying geocaching: cooperation vs conflict*, and *gamifying active aging: empowerment vs assistance*. The categories have been illustrated in the following sections. For privacy reasons, we used fictitious names to report participant quotes.

Gamifying Self-Tracking: From Motivation to Competition

A total of 3 factors motivated the elderly to participate in the project: curiosity about technology, the possibility of doing physical exercise with others, and an interest in cultural and historical aspects. From the analysis of the focus group conversations, the first aspect stands out: most of the participants stated that they were stimulated and intrigued by "using technology to support physical activity." The pedometer was found to be the most motivating element for the entire duration of the project, among the technologies proposed (ie, smartphones with the C: geo app and a pedometer bracelet):

> *I kept checking how many steps I had taken and tried to increase them...when you can check and you realise that at some point you are almost there, then you go for a walk* [Anna, participant]
>
> *After having exercised regularly for a month, activity becomes a physical need* [Gina, participant]
>
> *Normally when you walk you don't have the perception of taking so many steps* [Sara, participant]
>
> *You think you take so many steps at home...I never sit down...I take steps, but taking ten thousand...that's a long way!* [Diana, participant]

The use of the pedometer and thus the possibility of keeping track of the steps taken were, therefore, critical to motivating the participants to carry out physical activity and to making them aware of their physical activity level. Between the lines, we could also read the *coercive* aspect of the pedometer, which for some has become an almost necessary control tool. This trend was confirmed by the fact that, although the project was finished, most of the focus group participants continued to wear and use the bracelets. It was apparent that this coercive aspect had positive effects on both physical well-being and personal satisfaction:

> *When you see that you have walked that much, you are happy* [Gina, participant]

...and you have a result, because I feel better [Sara, participant]

The possibility of keeping track of the steps had positive effects on the motivation of the individual participants to obtain results. In addition, the awareness that their steps would influence the score of their team transformed the pedometer into an instrument of competition and comparison with other participants:

> *Initially it was like a game, but then came the rivalry...when I saw that we were below the score, I started to do some serious walking...at ten o'clock in the evening I was around increasing our step levels...I had a share of score that I could not fail to achieve...I had to stay on that level...so it was challenging* [Marina, participant]

The expectations regarding the team scores were high. The meetings before the hunts turned into opportunities for scores to be compared, and taking steps was seen almost as an obligation: a participant who missed a day's worth of walking to be present at a tournament of burraco (a card game) was (somewhat jokingly) accused of making their team lose. The creators of the hunt, despite not knowing which team they were assigned to, experienced this sense of challenge:

> *We didn't even know which team we were helping, so it was a matter of constancy...then it almost became an obsession...I would check and say to myself "I have to reach ten thousand, otherwise they'll hang me"* [Rita, creator]

The collective contribution to the group's score led to a strong team spirit, which heightened the competition during the hunts, as can be seen from the following discussion:

> *For example, when we were solving a riddle, the others would come and we looked in the other direction to give them wrong clues...We didn't even say goodbye! That's how much we felt we were enemies* [Fiona, participant]
>
> *I think the other team envied us...There was also some tipping off* [Sara, participant]

This strong sense of competition also emerged from the interaction between the focus group participants: opportunities for jokes, tensions, and even controversy between the members of the 2 teams were frequent.

Gamifying Geocaching: From Cooperation to Conflict

As mentioned earlier, the possibility of counting the steps and contributing to the team's score led to a strong sense of competition among the participants in the project. This sense of competition influenced the relations between individual participants and between the 2 teams and also involved other aspects of the gaming experience. All the focus group participants said that they were in a hurry to find all the caches:

> *We were mad...so we said "no comment"* [Alda, participant]
>
> *Also because often the riddle was not clear, but it needed to be understood* [Fiona, participant]
>
> *Did you focus on riddles or on the steps? Were we supposed to focus on taking many steps? As far as that is concerned, we succeeded...For the rest, I remember the hurry, but less hurry would have been nice for me since I don't know Trento...when the two teams met, people who knew each other didn't even say goodbye* [Marina, participant]
>
> *Also, about entering our observations...how were we supposed to write something?...it would have been nice to be able to write down some nice comments* [Margherita, participant]

Owing to the rush to reach the points of interest before the opposing team, some peculiar features of the geocaching game slipped into the background, such as carefully reading the cache description, leaving a comment after finding the cache, or adding photographs. Interviews with operators during outings confirm this trend: "Due to the rush to move on fast, we did not focus on other smartphone features such as selfies or writing sentences" (Mario, social operator). The sense of competition led to the rush to participate in the game, to solve the riddles, and to comment below the found caches, and this led to a deviation from the initial motivation that was the interest in historical and cultural aspects. However, this did not happen in the case of the creators, who, having not experienced the sense of team competition, maintained their enthusiasm for these aspects:

> *Even now when I walk around the city I look up...it is actually by looking up that you discover things...I discovered, on a building, a coat of arms that I didn't even know existed* [Gina, creator]

The lack of interest in some features of the game should not be attributed solely to the sense of competition. In fact, some features of the technologies contributed to this situation: first, the decision to entrust the team leader with the map with the smartphone led to a series of difficulties:

> *It was backlit, too shiny, and it was sunny...it was difficult to mark the riddles solved...if there were a bigger screen, everyone would see it...if you could see the whole journey perfectly...if you zoom in, you lose the other piece* [Alda, participant]
>
> *One of the important points was to read the description of the hunt stop...I read it aloud during the journey because there was no time...but of course there was the one with the stick behind you who couldn't hear* [Margherita, participant]

Hurdles were caused by the intrinsic features of the technology: the glare on the smartphone screen in the bright sun and a screen too small to display the whole map discouraged participants from using the tools necessary for carrying out the activity.

The comments reported so far have pointed out the ways in which some features of geocaching and aspects of gamification came into conflict, leading to hurdles that can become important when it comes to involving people of a certain age, as a participant stated in a very appropriate fashion:

> *If you want more people to have this chance, you should allow for much more time, including the chance to use the restroom, the chance to rest for those who walk with a stick...and to rest twice for those who feel they have walked too much...if instead you want to focus only on the treasure, then it is no longer for us* [Margherita, participant]

Gamifying Active Aging: From Empowerment to Assistance

The development of the project saw the constant presence of the center operators, from the first presentations to the hunt coplanning with the creators, the treasure hunt, and the final awarding. During all these phases, their experiential contribution to the user experience proved critical, both in terms of organization and support with the use of technologies as well as the management of conflicts among participants. The organizational context proved essential for the involvement of the participants. On the one hand, the fact that the senior center operators were already accustomed to proposing activities of this kind and paying attention to technology and active aging translated into having a pool of people already interested in the subject and, above all, people who trusted the operators:

> *Ease of involvement: we know who comes here for technology-related concerns, about eighty people* [Mario, social operator]
>
> *Gina, who hadn't been to the centre for a long time, was invited by Agnese, and came because she believed in it, trusted her* [Cristina, social operator]

On the other hand, the peculiarity of the center, owing to its recreational activities, allows the development of a strong network of relationships that, in turn, helps spread the word among people and facilitates chain involvement, thus allowing people to overcome the possible distrust of technologies:

> *The first time I didn't know what it was, then you told me, "do you want to participate?"* [Anna, participant]
>
> *Many, when they saw that I was enthusiastic, wanted to participate* [Rosa, participant]

The intervention of the operators was critical to the success of the project, particularly when it came to supporting the elderly in the use of technologies:

> *The first hunt was chaotic...we struggled because even the two team leaders felt awkward...we [operators] also had to help a lot with the phones...but in the end we learned by doing.* [Cristina, social operator]
>
> *We were always supported by Mario and Cristina, because if it happened they would jam.* [Marina, participant]
>
> *The team members who didn't have a smartphone weren't interested in using it. Maybe this fact may have taken the responsibility off the others with the risk that they would just follow...So I would say: "Come on, come here, let's look together where the points are" and there my role was that of regrouping them.* [Cristina, social operator]

The operators were able to support the participants in the use of the app, and their intervention proved to be decisive in directing the group's attention to the technology. This curbed an issue caused by the intrinsic characteristics of the technology and the spirit of competition among the participants.

Finally, the intervention of the operators allowed them to mediate some relational issues that emerged during the course of the project, such as the strong spirit of competition:

> *We had to work a lot on the team: we would encourage them to come closer, listen, help find the solution, look carefully...team members were granted a good level of autonomy but we had to provide assistance anyway as finding an agreement was not so simple.* [Mario, social operator]

As pointed out by Mario, the operator, although his main role was to support independent action on the part of the participants, in some cases, his presence was considered almost an excuse to delegate some activities: "No, we didn't take pictures, we were in a hurry...there was either Mario or Cristina [social operators] to do that" (Rosa, participant).

Discussion

Principal Findings

To the best of our knowledge, this is the first study on active aging that has combined elements of self-tracking and gamification with a geocaching activity. The success of the project, backed by the data on daily walk constancy of the participants and their satisfaction, is in line with the results of the studies that show how adding gamification to self-tracking can be a valid strategy to promote physical exercise among the elderly. However, through our study, we wanted to describe how the introduction of gamification elements required the inclusion of heterogeneous contextual and organizational elements that proved to be critical in determining the success of the project. Therefore, the study aimed to highlight how promoting active aging through technologies does not happen solely with the use of these tools but instead requires a wider process that involves a complex, articulated network.

Technologies as Responses to the Limits?

A study by Vartiainen and Tuunanen [48] pointed out the importance of the *hedonistic* value of geocaching: instead of stimulating the participants' sense of competition, the primary value remains to be the stronger sense of community that is aroused in geocachers. Our study aimed to propose a geocaching activity to promote active aging, integrated with elements of quantification (step counting) and gamification (team challenge). These elements—the pedometer bracelet, the team scoring mechanism, and the counting of the caches—proved to be highly motivating for the participants; however, they also led to a strong sense of competition, altering the nature of the activity and coming into conflict with some of the intrinsic characteristics of the game. For instance, in an endeavor to only come first, the participants did not use some features of the geocaching app, such as the ability to read cache descriptions, leave comments, and take photos. The haste and competition also led to some moments of tension and conflict among the participants. The correct use of the app and the mediation of conflicts could only take place thanks to the support of the center operators who mediated the conflicts and introduced older people to the use of technology. The technological tools were therefore integrated into a context that determined the methods of use: the familiarity of the social center with the technologies and the support of the social operators contributed to recreating an experience that was tailored to the skills and difficulties of the participants.

One of the most positive aspects of the project turned out to be the successful involvement and motivation of the participants. Although the involvement of users has become necessary for many technological innovations in different sectors [49], it has so far received little attention as an epistemic process itself. In most studies on active aging, there is an “accommodationist view” [50], which sees technologies designed as mere solutions to age-related needs and limitations. In our study, we wanted to go beyond this vision, focusing on the active involvement of the participants in every planning phase, from the conception of the paths to the presentation of the project to their peers. We wished to create a protagonism of elderly people, who do not feel abandoned but are accompanied by the caregivers on their journey toward autonomy and independence. This path is not exempt from contradictions, as is evident from the fact that the intervention of operators can also become an excuse to delegate the use of technology. However, this confirms that these limits derive from the interactions among players, institutions, and material objects [51]. As we highlighted in this paper, the introduction of technologies for active aging inevitably involves an interactive and evolving commitment with the social and cultural spheres of the users with all their difficulties, abilities, and constant reinventions.

Study Limitations and Future Developments

Our study had some limitations and some lessons learned. The first limitation of the study concerned the limited nature of the sample examined. Although this was the first pilot study in which the qualitative methodologies proved to be most appropriate in achieving the research objectives, the project nevertheless took the first step toward the expansion to other organizations in the local area (such as senior centers, clubs, and schools), which would provide a larger sample. To this end, in the year following the pilot study described in this paper, the Impronte project saw 2 further additions: the first involved a school, in addition to the senior center, with a view to intergenerational exchange, and the second saw the inclusion of a natural park, with the objective of local seniors devising routes for tourists. The idea was to continue on this track, owing to the development of a geocaching platform designed by and for the elderly, which would allow heterogeneous and, in some cases, isolated contexts to be reached. A second limitation—and a lesson learned—of the study was the lack of a collection of pre- and postactivity data: analyzing the amount of steps taken by older people before and after participating in the project could provide interesting statistics on the motivation to maintain healthy lifestyles. Another lesson learned is with regard to encouraging participants to learn to use technology through geocaching. Our study provided that only 1 team component managed the phone during the game. This is because we had to decide whether, on the one hand, to encourage participants to learn to use technology, but with the risk that each participant would have to participate on his or her own and be isolated from the rest of the group, or whether, on the other hand, to encourage the social aspect of the game, stimulating the communication between participants. We chose the second option, but at the expense of encouragement to learn to use technology. It would be interesting to investigate if providing all participants with the opportunity to manage a phone would encourage them to learn to use technology.

Conclusions

Activities in favor of active aging that involve the use of new technologies can foster the participation of elderly people, as was evident in the success of this project: an activity that allows outdoor exercise through a playful activity such as geocaching and the use of self-tracking tools can encourage and motivate older people to perform physical activity, as shown by the quantitative data. However, this paper aimed to show how technology use of older people was reshaped and contextualized during the project, with some aspects of the activity prioritized and others sent to the background.

It is important to go beyond the fact that users simply have to be involved and that the involvement of multiple users automatically leads to better knowledge. We propose that it is essential to develop a detailed understanding of how, when, and why they should be involved and also when and how they should not. Above all, users should actively participate in the creation of project contents from a peer-to-peer perspective, as was the case for the project described in this paper.

This approach can enable stakeholders, policy makers, and innovators to better understand the future of aging and address the challenges of innovation for older users more completely. Going beyond an *accommodating* vision that sees science and technology as mere solutions is critical to designing technologies that are functional for seniors (and their caregivers) and allow them to embrace the challenge of active aging.

Acknowledgments

The Impronte project is a part of a series of initiatives aimed at promoting healthy lifestyles by the TrentinoSalute 4.0 Competence Center, which includes the Department of Health and Social Solidarity of the Province of Trento, Italy; the Provincial Health Services; and the Bruno Kessler Foundation. The study involved the *Contrada Larga* senior center of Trento, managed by the Kaleidoscopio social cooperative. The authors would like to thank the social operators and all the elderly involved in this study.

Conflicts of Interest

None declared.

References

1. Delai N. Seniors & Continuity of Care. Identify a Shared Medium-Term Strategy for the Conditions of Non-Autonomy. Book in Italian. Anziani & continuità assistenziale: individuare una strategia condivisa di medio periodo per le condizioni della non autonomia. Italy: Franco Angeli; 2013.
2. WHO. Global Recommendations on Physical Activity for Health. Geneva: World Health Organization; 2010.
3. Wen CP, Wai JP, Tsai MK, Yang YC, Cheng TY, Lee MC, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet 2011 Oct 1;378(9798):1244-1253. [doi: 10.1016/S0140-6736(11)60749-6] [Medline: 21846575]
4. Vollenbroek-Hutten M, Pais S, Ponce S, Dekker-van Weering M, Jansen-Kosterink S, Schena F, et al. Rest Rust ! Physical active for active and healthy ageing. Transl Med UniSa 2015 Dec;13:19-28 [FREE Full text] [Medline: 27042429]
5. Czaja SJ, Lee CC. The impact of aging on access to technology. Univ Access Inf Soc 2007;5(4):341-349. [doi: 10.1007/s10209-006-0060-x]
6. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78 [FREE Full text] [doi: 10.1177/0141076813480996] [Medline: 23481424]
7. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
8. Silveira P, van de Langenberg R, Reve EH, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: a phase II preclinical exploratory trial. J Med Internet Res 2013 Aug 12;15(8):e159 [FREE Full text] [doi: 10.2196/jmir.2579] [Medline: 23939401]
9. Garcia JA, Navarro KF, Schoene D, Smith ST, Pisan Y. Exergames for the elderly: towards an embedded Kinect-based clinical test of falls risk. Stud Health Technol Inform 2012;178:51-57. [Medline: 22797019]
10. Garcia JA, Navarro K, Lawrence E. Serious Games to Improve the Physical Health of the Elderly: A Categorization Scheme. In: Proceedings of The Fourth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services. 2011 Presented at: CENTRIC'11; October 23-29, 2011; Barcelona, Spain URL: https://www.researchgate.net/publication/244993334_Serious_games_to_improve_the_physical_health_of_the_elderly_a_categorization_scheme
11. Boletsis C, McCallum S. Smartkuber: A serious game for cognitive health screening of elderly players. Games Health J 2016 Aug;5(4):241-251. [doi: 10.1089/g4h.2015.0107] [Medline: 27192473]
12. Makizako H, Shimada H, Doi T, Tsutsumimoto K, Lee S, Hotta R, et al. Cognitive functioning and walking speed in older adults as predictors of limitations in self-reported instrumental activity of daily living: prospective findings from the Obu Study of Health Promotion for the Elderly. Int J Environ Res Public Health 2015 Mar 11;12(3):3002-3013 [FREE Full text] [doi: 10.3390/ijerph120303002] [Medline: 25768239]
13. Purpura S, Schwanda V, Williams K, Stubler W, Sengers P. Fit4life: The Design of a Persuasive Technology Promoting Healthy Behavior and Ideal Weight. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7 - 12, 2011; Vancouver, Canada p. 423-432. [doi: 10.1145/1978942.1979003]
14. Thieme A, Comber R, Miebach J, Weeden J, Kraemer N, Lawson S, et al. 'We've Bin Watching You': Designing for Reflection and Social Persuasion to Promote Sustainable Lifestyles. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 5 - 10, 2012; Austin, Texas p. 2337-2346. [doi: 10.1145/2207676.2208394]
15. Tudor-Locke C. Education Resources Information Center. 2002 Jun. Taking Steps toward Increased Physical Activity: Using Pedometers To Measure and Motivate URL: https://eric.ed.gov/?id=ED470689 [accessed 2020-03-23]
16. Cameirão M, Bermúdez S, Verschure P. Virtual reality based upper extremity rehabilitation following stroke: a review. J Cyber Ther Rehabil 2008;1(1):63-74 [FREE Full text]
17. Bernd L. Localytics. 2011 Jan 31. 26% of Mobile Apps Downloaded in 2010 Were Used Just Once URL: http://info.localytics.com/blog/first-impressions-26-percent-of-apps-downloaded-used-just-once [accessed 2020-03-06]
18. Fankhauser D. Mashable. 2013 May 17. Is Gamification Just a Fad? URL: http://mashable.com/2013/05/17/gamification-buzzword/ [accessed 2014-03-14]
19. Ramachandran V. Mashable. 2013 Aug 3. AchieveMint Gives You Cash for Doing Healthy Activities URL: https://mashable.com/2013/08/02/achievemint/ [accessed 2014-03-14]
20. Young H, Miyamoto S, Ward D, Dharmar M, Tang-Feldman Y, Berglund L. Sustained effects of a nurse coaching intervention via telehealth to improve health behavior change in diabetes. Telemed J E Health 2014 Sep;20(9):828-834 [FREE Full text] [doi: 10.1089/tmj.2013.0326] [Medline: 25061688]
21. de Vette F, Tabak M, Dekker-van Weering M, Vollenbroek-Hutten M. Engaging elderly people in telemedicine through gamification. JMIR Serious Games 2015 Dec 18;3(2):e9 [FREE Full text] [doi: 10.2196/games.4561] [Medline: 26685287]

22. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
23. Minge M, Bürglen J, Cymek D. Exploring the Potential of Gameful Interaction Design of ICT for the Elderly. In: Proceedings of the International Conference on Human-Computer Interaction. 2014 Presented at: HCI'14; June 22-27, 2014; Crete, Greece p. 304-309. [doi: 10.1007/978-3-319-07854-0_54]
24. Dionigi RA, Horton S, Baker J. Negotiations of the ageing process: older adults' stories of sports participation. Sport Educ Soc 2013;18(3):370-387. [doi: 10.1080/13573322.2011.589832]
25. Lomborg S, Frandsen K. Self-tracking as communication. Inf Commun Soc 2016;19(7):1015-1027. [doi: 10.1080/1369118x.2015.1067710]
26. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
27. Spillers F, Asimakopoulos S. Does Social User Experience Improve Motivation for Runners? In: Proceedings of the International Conference of Design, User Experience, and Usability. 2014 Presented at: DUXU'14; June 22-27, 2014; Heraklion, Crete, Greece p. 358-369. [doi: 10.1007/978-3-319-07638-6_35]
28. Tauer JM, Harackiewicz JM. The effects of cooperation and competition on intrinsic motivation and performance. J Pers Soc Psychol 2004 Jun;86(6):849-861. [doi: 10.1037/0022-3514.86.6.849] [Medline: 15149259]
29. Rapp A, Cena F. Affordances for Self-Tracking Wearable Devices. In: Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015 Presented at: ISWC'15; September 7 - 11, 2015; Osaka, Japan p. 141-142. [doi: 10.1145/2802083.2802090]
30. Sherman E. Geocaching: Hike And Seek With Your Gps. New York: Apress; 2004.
31. Cattan M, White M, Bond J, Learmouth A. Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions. Aging Soc 2005 Jan;25(1):41-67. [doi: 10.1017/S0144686X04002594]
32. Masi CM, Chen H, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev 2011 Aug;15(3):219-266. [doi: 10.1177/1088868310377394] [Medline: 20716644]
33. Hooper CJ, Rettberg JW. Experiences with Geographical Collaborative Systems: Playfulness in Geosocial Networks and Geocaching. In: Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2011 Presented at: MobileHCI'11; August 30–September 2, 2011; Stockholm, Sweden.
34. Cameron L. Geocaching Handbook: The Guide for Family Friendly, High-Tech Treasure Hunting. Lanham, MD: Rowman & Littlefield; 2017.
35. Taylor JK, Kremer D, Pebworth K, Werner P. Geocaching for Schools and Communities. Champaign, IL: Human Kinetics; 2010.
36. Gram-Hansen LB. Geocaching in a Persuasive Perspective. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive'09; April 26 - 29, 2009; California, Claremont, USA p. 1-8. [doi: 10.1145/1541948.1541993]
37. Goffman E. On fieldwork. J Contemp Ethnogr 1989 Jul;18(2):123-132. [doi: 10.1177/089124189018002001]
38. Frisina A. Focus Group. A Practical Guide. Book in Italian. Focus group: una guida pratica. Italy: Il mulino; 2010.
39. Liamputtong P. Focus Group Methodology: Principles And Practice. London, UK: Sage Publications, Ltd; 2011.
40. Heinz M, Martin P, Margrett JA, Yearns M, Franke W, Yang H, et al. Perceptions of technology among older adults. J Gerontol Nurs 2013 Jan;39(1):42-51. [doi: 10.3928/00989134-20121204-04] [Medline: 23244061]
41. Morgan D. The Focus Group Guidebook. Thousand Oaks, CA: Sage Publications, Inc; 1998.
42. Altheide DL. Reflections: Ethnographic content analysis. Qual Sociol 1987;10(1):65-77. [doi: 10.1007/bf00988269]
43. Morgan DL. Qualitative content analysis: a guide to paths not taken. Qual Health Res 1993 Feb;3(1):112-121. [doi: 10.1177/104973239300300107] [Medline: 8457790]
44. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
45. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
46. Zhang Y, Wildemuth BM. UT iSchool | The University of Texas at Austin. 2009. Qualitative Analysis of Content URL: https://www.ischool.utexas.edu/~yanz/Content_analysis.pdf [accessed 2020-03-23]
47. Vicsek L. A scheme for analyzing the results of focus groups. Int J Qual Methods 2007 Dec;6(4):20-34. [doi: 10.1177/160940690700600402]
48. Vartiainen T, Tuunanen T. Contradictions in Value Co-Creation in IT-Enabled Services: A Case of Geocaching. In: Proceedings of the 25th Australasian Conference on Information Systems. 2014 Dec 10 Presented at: ACIS'14; December 8-10, 2014; Auckland, New Zealand URL: https://core.ac.uk/download/pdf/56364778.pdf
49. Peace S, Hughes J. Reflecting on User Involvement and Participatory Research. London, UK: Centre for Policy on Ageing; 2010.
50. Dannefer D. Age, the life course, and the sociological imagination: Prospects for theory. In: Binstock RH, George LK, editors. Handbook of Aging and the Social Sciences. Cambridge, Massachusetts: Academic Press; 2011:3-16.

51. Peine A, Herrmann AM. The sources of use knowledge: Towards integrating the dynamics of technology use and design in the articulation of societal challenges. Technol Forecast Soc Change 2012 Oct;79(8):1495-1512. [doi: 10.1016/j.techfore.2012.04.014]

Abbreviations

APIs: application programming interfaces

Edited by G Eysenbach; submitted 03.07.19; peer-reviewed by M Baez, A Mahnke, E Rubenstein; comments to author 21.10.19; revised version received 30.01.20; accepted 22.02.20; published 11.06.20.

Please cite as:
Fornasini S, Dianti M, Bacchiega A, Forti S, Conforti D
Using Geocaching to Promote Active Aging: Qualitative Study
J Med Internet Res 2020;22(6):e15339
URL: http://www.jmir.org/2020/6/e15339/
doi:10.2196/15339
PMID:32525486

Original Paper

Understanding Weight Loss via Online Discussions: Content Analysis of Reddit Posts Using Topic Modeling and Word Clustering Techniques

Yang Liu[1], MSc; Zhijun Yin[2,3], PhD

[1]College of Computer Science and Technology, Changchun Normal University, Changchun, China
[2]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
[3]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, United States

Corresponding Author:
Zhijun Yin, PhD
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Ave.
14th floor, Suite 1475
Nashville, TN, 37023
United States
Phone: 1 6159363690
Email: zhijun.yin@vanderbilt.edu

Abstract

Background: Maintaining a healthy weight can reduce the risk of developing many diseases, including type 2 diabetes, hypertension, and certain types of cancers. Online social media platforms are popular among people seeking social support regarding weight loss and sharing their weight loss experiences, which provides opportunities for learning about weight loss behaviors.

Objective: This study aimed to investigate the extent to which the content posted by users in the *r/loseit* subreddit, an online community for discussing weight loss, and online interactions were associated with their weight loss in terms of the number of replies and votes that these users received.

Methods: All posts that were published before January 2018 in r/loseit were collected. We focused on users who revealed their start weight, current weight, and goal weight and were active in this online community for at least 30 days. A topic modeling technique and a hierarchical clustering algorithm were used to obtain both global topics and local word semantic clusters. Finally, we used a regression model to learn the association between weight loss and topics, word semantic clusters, and online interactions.

Results: Our data comprised 477,904 posts that were published by 7660 users within a span of 7 years. We identified 25 topics, including food and drinks, calories, exercises, family members and friends, and communication. Our results showed that the start weight (β=.823; $P<.001$), active days (β=.017; P=.009), and median number of votes (β=.263; P=.02), mentions of exercises (β=.145; $P<.001$), and nutrition (β=.120; $P<.001$) were associated with higher weight loss. Users who lost more weight might be motivated by the negative emotions (β=−.098; $P<.001$) that they experienced before starting the journey of weight loss. In contrast, users who mentioned vacations (β=−.108; P=.005) and payments (β=−.112; P=.001) tended to experience relatively less weight loss. Mentions of family members (β=−.031; P=.03) and employment status (β=−.041; P=.03) were associated with less weight loss as well.

Conclusions: Our study showed that both online interactions and offline activities were associated with weight loss, suggesting that future interventions based on existing online platforms should focus on both aspects. Our findings suggest that online personal health data can be used to learn about health-related behaviors effectively.

(J Med Internet Res 2020;22(6):e13745) doi:10.2196/13745

KEYWORDS

weight loss; online health community; machine learning; topic modeling; word2vec; hierarchical clustering; consumer health

Introduction

Background

Maintaining a healthy weight can reduce the risk of developing many diseases, including type 2 diabetes, hypertension, heart disease and strokes, kidney diseases, and certain types of cancers [1-6]. Unfortunately, overweight or obesity has nowadays become a public health crisis that impacts many Americans. For example, it was reported that nearly 94 million US adults were affected by obesity in 2015 and the annual medical cost was approximately US $150 billion [7]. To promote public health and help control overweight and obesity, it is critical to understand what factors are associated with weight loss and design effective weight loss interventions.

Over the past decade, people have been increasingly leveraging online social media platforms to share personal experiences and seek social support regarding weight loss, understand the impact of obesity, and learn their interpretation as contributing factors to a healthy life [8]. This huge amount of online information enables health care providers and researchers to gain insights into both public and personal health. For example, studies showed that what people shared on Instagram and Twitter could be used to effectively assess obesity prevalence in the United States [9,10]. While a content analysis showed that Twitter users were more likely to discuss weight loss during and after holidays [11], a survey-based study suggested that the interactions on Twitter were too brief and shallow [12], which might constrain its users to gain deep social support. This is, however, a major motivation for people to engage in other online weight loss forums or communities [13].

As such, recent studies, which aimed to prevent obesity and promote healthy weight loss, tended to incorporate online social media platforms into their intervention design, but the effects were found to be mixed. While some investigations showed that these online platforms had the potential for an innovative weight loss intervention [14,15], others found their effects were limited because of a low retention and engagement rate [16,17]. Moreover, a meta-analysis of over 2000 studies concluded that the effects of the interventions incorporating online social networks were very modest in improving health-promoting behaviors [18]. This suggested that before designing interventions based on qualitative evidence that online support and engagement are helpful [19], a quantitative analysis is necessary in determining how online interactions with other users (as a potential external influential factor), and a user's offline activities recorded in online discussions (as a potential internal driving factor), are associated with their weight loss.

Current Research and Its Limitation

It should be noted that there were studies showing that consistent online activities (eg, updating progress in weight loss and interacting with others) in online weight loss programs or training were associated with higher weight loss [20-22]. Particularly, a study using data from *r/loseit* showed that higher BMI levels and higher online activities were associated with more weight loss [23]. Similarly, another study based on causal inference found that, on average, users who received comments in *r/loseit* lost 9 lb more than users who did not receive any comments [24]. Although online interactions were shown to have a significant impact on weight loss [25], existing studies focused less on the content of posts. While both the aforementioned studies used topic modeling to extract topics, they merely focused on the most popular ones. This method, however, was too general to identify detailed offline activities that were disclosed in such a large number of posts. Moreover, these studies used the data that were generated during 2010 and 2014, which, as we showed later, consisted of only a small fraction of the posts published in *r/loseit*.

In fact, highlighting both online interactions and personal offline activities aligns with social cognitive theory (SCT) [26]. The theory emphasizes that external and internal social reinforcement together lead to behavior change in a dynamic fashion and is often applied to guild the design of effective intervention strategies [27]. This suggests that focusing on either online interactions or personal offline activities but not both might lead to an incomplete view of the roles of online communities in the process of weight loss, but this needs to be examined and confirmed with evidence.

Objective

Therefore, considering the limitation of previous studies and inspired by SCT, we aimed to investigate the extent to which the offline activities communicated by users in the *r/loseit* subreddit, an online community for discussing weight loss, and online interactions were associated with their weight loss. Specially, we focused on a data set consisting of 477,904 posts that were published by 7660 users before January 2018 in *r/loseit*. We used the self-reported weight change to measure weight loss and the average number of comments and votes that they received from other users to characterize online interactions. We applied both topic modeling and word clustering to obtain detailed and interpretable contributing factors from online posts. Finally, we used a linear regression model to quantitively examine the association between online interactions, factors described in online discussions, and the amount of weight loss.

Our work provided evidence that an online social media platform can serve as an effective data source to understand weight loss, and our findings implied that in future weight loss analyses or interventions, online interactions should be considered as a factor that influences long-term self-efficacy.

Methods

Data

Our data were collected from *r/loseit*, a subreddit focusing on weight loss in Reddit, an online discussion platform. Within the subreddit, users can either publish a *submission* to start a new discussion thread or make comments on either a submission or another comment to an existing discussion thread. For simplicity, we used the word *post* to denote either a submission or a comment when we did not differentiate them. In addition, Reddit users can upvote or downvote a comment but can only upvote a submission. Furthermore, users in many subreddits are allowed to enter text or symbols into a *flare*, which appears next to their usernames in a post, to show some basic information of themselves. For example, in *r/loseit*, users can show their *start weight*, *current weight*, and *goal weight* and even their *gender*, *age*, and *height* information in flairs. However, as creating a *flare* is not required, users can ignore it when publishing a post.

In this study, we used the Python Reddit API Wrapper python package (version 5.3.0) to extract data from the Reddit application programming interface. Specifically, we collected all the posts in *r/loseit* that were published before January 13, 2018. We used the flares tagged with usernames to confine our study cohorts to those users who disclosed their start, current, and goal weights and were active for at least 30 days in this subreddit [23]. It should be noted that we did not ask for permission to use the data from the Reddit community because the data are publicly accessible. However, we never tried to identify any Reddit user by linking their Reddit data with additional data sets. All the results, and post samples presented in this paper, were carefully examined and revised such that no personally identifiable information was disclosed.

Topic Modeling and Word Semantic Clustering

Owing to high dimensionality, noise, and ambiguity of natural language text, processing and analyzing raw post content are often challenging, and the analyzed results are difficult to interpret. In this study, we used 2 types of methods to mitigate this problem: topic modeling and word semantic clustering based on low dimensional representation (eg, word2vec) [28]. While topic modeling can help identify themes in a global context, word semantic clusters can provide more detailed concepts in a local context [29].

Specifically, we used the implementation of latent Dirichlet allocation (LDA) in Mallet (version 2.0.8) to identify the main topics of online discussions in *r/loseit* [30]. Since LDA is an unsupervised algorithm, we relied on a coherence score to determine the optimal number of topics [31]. In this study, we trained LDA models for 5 to 75 topics (with a step size of 5) on all of the posts and chose the number of topics that was corresponding to the highest coherence score. To mitigate word sparsity, we only kept nouns, verbs, adjectives, and adverbs.

To obtain word semantic clusters, we relied on the Google pretrained word2vec model because our data set was not large enough to train an accurate word2vec model. We relied on the standard deviation of cluster size to determine the optimal number of clusters [29]. Specifically, we used a hierarchical clustering algorithm with 25 to 1000 clusters (with a step size of 25) and used the elbow rule to the standard deviation of the number of words in clusters. Intuitively, a large word cluster is more likely to contain multiple concepts, while a small word cluster is more likely to have little contribution to reducing hundreds of thousands of word dimensions [29].

Regression Analysis

In this study, we investigated the association between weight loss and online discussions by using a linear regression model. Specifically, we characterized a user's online discussion by using the following predictors:

- The days that the user was active in the subreddit.
- The number of posts that the user published.
- The topics conveyed in the posts, measured by topic distribution.
- The word semantic clusters, measured by term frequency–inverse document frequency values.
- The median karma score or votes that the user received for each post, measured by subtracting the number of upvotes from the number of downvotes [32].
- The median number of comments for each post that the user published.

We used weight loss, measured by subtracting the start weight from the current weight, as the outcome variable of the regression model. As the distribution of the weight loss variable is right-skewed, we log-transformed it before feeding it into the regression model. All the predictors were normalized and scaled into a range of (0, 1). It should be noted that the active days and the number of comments were also log-transformed because of their right-skewed distributions.

Considering that a person who had higher weight at the beginning is more likely to lose more weight, we also introduced *start weight* as a control variable in the model. Before applying the regression analysis, we used the *findCorrelation* function, as implemented in the caret R package (version 6.0-81), with a cutoff of 0.3 to remove correlated predictors. We reported predictors with a statistical significance level of .05.

Results

Data Statistics

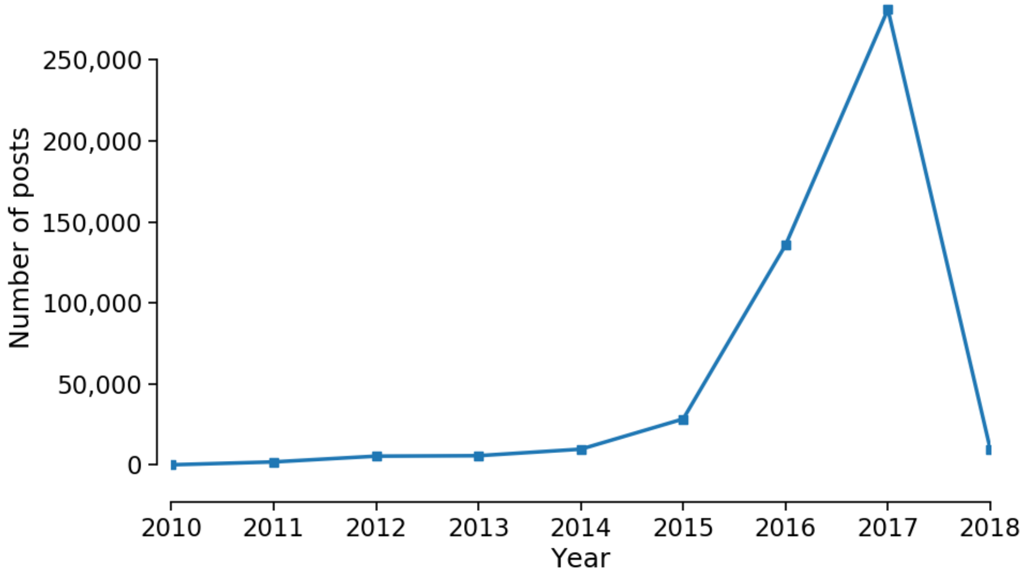
We collected 2,526,277 posts published by 205,722 users during the period between July 30, 2010, and January 13, 2018. Focusing on the users who disclosed their start, current, and goal weights in flairs, we finally obtained 7660 users with a total of 477,904 posts. These posts included 16,332 submissions and 461,572 comments. Table 1 summarizes the basic statistics of key factors regarding this study cohort. From the table, we observed that most posts received a small number of comments and karma scores.

Figure 1 shows the number of posts for each year during 2010 and 2018. From the figure, we can see that the number of posts surged after 2014. It should be noted that the number of posts in 2018 was small because there were only 2-week data before we stopped data collection.

Table 1. Basic statistics of the study cohort.

Statistics	25th percentile	Median	75th percentile
The number of posts per user, n	8	19	52
The number of words per post, n	18	39	78
The start weight (lb)	183.0	220.0	265.0
Weight loss (lb)	16.0	30.0	52.0
Active days, n	21	89	281
The number of comments for each post, n	1	1	1
The karma score for each post, n	1	2	3

Figure 1. The number of posts published in this subreddit in each year (from July 30, 2010, to January 13, 2018).



Topics Discovered in Online Discussions

We identified 25 topics that were corresponding to the highest coherence score (Multimedia Appendix 1). Table 2 shows the inferred topics, their marginal distribution, and the most relevant terms. The marginal distribution of a topic was measured by the probability that the topic was sampled from online discussions, while the relevance of a term was measured by the probability that it was sampled from a topic. The identifier of each topic was named based on the descending order of their topic distribution. For example, topic T1, talking about drinks, had the highest distribution, while topic T25, one of the weight change–related topics, had the lowest topic distribution.

We also manually summarized the 25 topics into 11 categories and provided the associated labels in Table 2. The table shows that people in this subreddit often talked about food and drinks, exercise, calorie, clothes, time, health issues, weight change, feelings, plans, and communication.

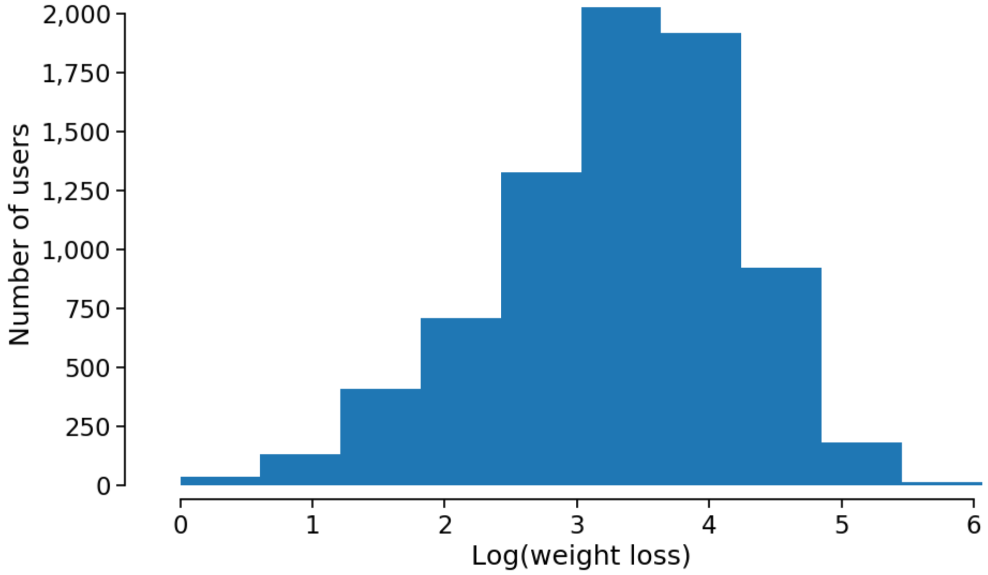
Table 2. The 25 topics that were identified from r/loseit. The sample terms were ordered based on their relevance to the topic. The distribution of each topic is calculated on all the posts published by these users.

Label and ID	Most relevant terms	Distribution (probability)
Food or drinks		
T2	water, drink, lot, drinking, make, soda, cut, diet, beer, bit, thing, alcohol	.053
T11	eat, food, meal, eating, lunch, dinner, hungry, calorie, breakfast, pizza	.051
T21	food, diet, eating, eat, low, protein, fat, calorie, lot, high, carbs, cico	.049
T22	veggie, chicken, salad, cheese, add, recipe, make, cook, egg, rice, meat	.048
Health issues		
T5	doctor, body, issue, problem, health, level, pain, energy, effect, surgery	.052
Family or friends		
T6	friend, family, guy, kid, mom, school, husband, life, live, told, girl	.052
Calorie		
T7	calorie, deficit, exercise, tdee, eat, counting, count, eating, daily, mfp	.052
T8	calorie, chocolate, sugar, sweet, snack, coffee, craving, bar, fruit, cal, cup	.051
Exercise		
T9	run, running, walk, minute, time, walking, mile, step, hour, half, long	.051
T16	gym, exercise, muscle, workout, start, working, work, body, cardio	.049
Weight change		
T3	lb, year, lost, month, started, pound, back, ago, gained, weight, starting	.052
T17	weight, lose, loss, fat, body, gain, healthy, bmi, height, normal, range	.049
T25	week, scale, weight, pound, number, plateau, drop, weighing, daily	.043
Feelings		
T12	thing, work, make, find, lot, time, hard, put, give, won, easy, worry	.051
T13	goal, hit, progress, end, picture, feel, close, happy, set, time, challenge	.050
T14	feel, feeling, eating, binge, bad, stop, time, control, hard, stress, struggle	.050
Communication or encouragement		
T1	good, pretty, yeah, luck, lol, idea, feel, kind, sound, guess, bit, haha, stuff	.053
T10	people, person, health, comment, understand, talk, mental, life, care	.051
T23	post, read, check, https_www, question, loseit, app, mfp, reddit, http	.048
T24	great, love, awesome, job, amazing, congrats, hope, sound, glad, similar	.046
Plan or decision		
T4	day, week, back, track, maintenance, couple, plan, log, logging, cheat	.052
T18	change, time, make, start, long, life, habit, making, thing, choice, healthy	.049
T20	time, thought, made, felt, wanted, started, back, needed, decided, found	.049
Clothes		
T19	fit, size, clothes, big, wear, buy, pant, bought, short, shirt, dress, store	.049
Time		
T15	today, morning, night, yesterday, work, tomorrow, weekend, day, home	.050

Regression Analysis

We chose 425 as the optimal number of word clusters based on the elbow rule (Multimedia Appendix 1). The word semantic clusters, together with other proposed predictors, were applied to fitting a linear regression model. After examining the feature correlation, we included 6 topics and 402 word semantic clusters into the regression model. Figure 2 shows the distribution of the log-transformed weight loss, which matches the loose constraint of applying a linear regression model.

Figure 2. Histogram of log-transformed weight loss.



Goodness of Fit and Non–content Predictors

The fitted linear regression model had an adjusted R^2=0.315, $F_{7,191}$=9.553, and P<.001, suggesting that the model with proposed predictors predicted weight loss better than the basic intercept-only model. Among the non–content-related predictors, the *start weight* (β=.823; *P*<.001), the *active days* in the subreddit (β=.02; *P*=.009), and the *median karma score* (β=.263; *P*=.02) were associated with higher weight loss. However, the *median number of comments* was not significantly associated with weight loss in our analysis (β=.001; *P*=.95).

Topic-Related Predictors

There were 3 topics that were found to be significantly associated with weight loss. Topics T16 (exercise in the gym) and T19 (purchase of clothes) were associated with higher weight loss (β=.072; *P*=.007 and β=.080; *P*<.001, respectively), while topic T7 (counting calorie) was associated with lower weight loss (β=−.074; *P*=.007).

Word Semantic Cluster–Related Predictors

Table 3 shows word semantic clusters that were significantly associated with higher weight loss, which are summarized below.

Table 3. Word semantic clusters that were statistically significant with respect to weight loss. Sample words are presented based on their distances to the center of a cluster in ascending order.

ID	Sample words	Beta	z score	SD	*P* value
C258	jacket, sweater, socks, sweaters, shirt, shirts, scarf, plaid, sweatshirts, jackets, coats, sandals, sweatshirt	.145	6.126	0.024	<.001
C383	emulate, replicate, magic, magical, superhuman, miraculous, recreate, miracle, feats, imitate, miracles	.130	4.699	0.028	<.001
C46	zelda, kinect, newb, wii, noob, mod, addons, nintendo, mmo, mario, xbox, mods, gamers, sims, addon, gamer	.129	1.988	0.065	.047
C93	cardio, aerobic, yoga, boxercise, gymming, pilates, plyometrics, treadmill, stairmaster, calisthenics	.128	4.567	0.028	<.001
C301	protein, metabolism, metabolic, leptin, lipids, proteins, serotonin, enzyme, enzymes, hormone, ghrelin	.120	5.143	0.023	<.001
C334	angst, frustration, anguish, anxiety, unhappiness, despair, bitterness, anger, loathing, resentment	.098	3.826	0.025	<.001
C304	musical, orchestra, opera, rehearsal, rehearsals, singers, performances, theater, choir, audition, artistic	.093	3.163	0.029	.002
C203	coach, powerlifting, teammates, captain, coached, squad, powerlifter, weightlifting, players, coaches	.085	2.027	0.042	.04
C110	younger, age, youngest, older, aged, ages, novice, old, beginners, beginner, oldest, lifelong, intermediate	.081	4.643	0.017	<.001
C24	straightforward, efficient, economical, adaptable, intuitive, inexpensive, sophisticated, expensive, easy	.081	4.405	0.018	<.001
C143	copper, zinc, moly, gold, ore, metals, tungsten, mineral, nickel, minerals, lithium, silver, metal, mines	.073	2.860	0.025	.004
C193	flower, flowers, bulb, dandelion, fragrant, roses, twigs, leaf, daffodils, buds, hibiscus, tree, crocus, trees, moss	.073	2.520	0.029	.01
C291	restaurant, restaurants, diner, deli, resturant, café, diners, bistro, cafe, pizzeria, steakhouse, resturants	.065	2.153	0.03	.03
C368	fitted, styled, proportioned, equipped, smooth, positioned, engineered, designed, smoother, molded	.059	2.584	0.023	.01
C293	discarded, trash, garbage, trashcan, bins, dumpster, recycle, dump, rubbish, scraps, leftovers, recycled, bin	.057	2.243	0.025	.03
C108	socialize, invite, catered, inviting, socializing, catering, cater, hosting, caters, hosted, invited, accommodating	.055	2.982	0.018	.003
C233	driveway, pavement, sidewalk, sidewalks, road, roads, highway, street, freeway, streets, alley, intersection	.054	2.062	0.026	.04
C66	cookouts, picnics, barbecues, cookout, feasts, picnic, potluck, togethers, barbeque, barbecue, potlucks, feast	.052	2.280	0.023	.02
C285	gerd, brussel, havn, ddp, utrecht, stephan, scandinavia, bulgarian, cico, sicilian	.045	2.797	0.016	.005
C25	gluttony, indulgence, vices, gluttonous, sinful, stupidity, hedonistic, sin, decadent, indulgent, excesses, sloth	.044	2.563	0.017	.01
C200	walkable, downtown, waterfront, neighborhoods, neighborhood, park, city, touristy, parks, suburbs	.041	1.967	0.021	.049

Exercise-Related Clusters

Exercise-related clusters included C258 (clothes), C46 (Xbox games), C93 (cardio), C203 (coaching), C233 (roads), C285 (city-related food, exercise), and C200 (hangouts places). The following are some examples that were composed by users with higher weight loss:

> *6 shirts, 2 sweaters, 6 jeans and my belts are on the last hole! I was so surprised when I read a post about new cloths and the belt was mentioned...I never knew I was already on the last hole D: I don’t need to use it, but I certainly can!*
>
> *You forget that with muscles, you'll burn more fat. 80% of my training consists of strength and I do some small amount of cardio on the side.*
>
> *I liked the suggestions and coaching from the jawbone app, but the overestimation is why I got a Fitbit with a heart rate monitor.*

Diet- or Dining-Related Clusters

Diet- or dining-related clusters included C301 (nutrition), C143 (minerals), C291 (restaurant), C293 (leftover), C108 (hosting), C66 (cookout), C25 (evil gluttony), and C285 (city-related food, exercise). The following are some examples that were communicated in related posts:

> *today I eat a pancake with plain greek yoghurt as a pre-workout in-work snack, after gym I will inhale two greek yoghurts with two spoons of protein powder, which catapults me to 170g of protein today.*
>
> *I think I've spent more time replacing bad items (like I have asparagus and Brussel sprouts instead of freezer fries now, for example) with better options than I have really giving things up.*
>
> *I learned the same thing with pizza. I love pizza, but nowadays I would much rather enjoy a slice of pizza from a good local restaurant, than an entire pizza from Domino's or Papa John's.*

> *And knowing myself before I started losing weight, if I can do it so can everyone else. I was the embodiment of laziness, gluttony and excuses.*

Other Clusters

Other clusters included C304 (performances), C334 (negative emotions), C24 (simple, straightforward, economic), C193 (flowers), C110 (beginner), and C383 (replicate). The following are some examples of these clusters:

> *I ended up just drinking water at the theater because everything was either full of calories or full of sugar. ended up having some leftover chicken and a banana when I got home.*
>
> *About the anxiety, yea same boat. I struggle with social anxiety since I'm 13 years old but it did get a lot better over the last few months.*
>
> *Look into a Hot Pot—they are relatively inexpensive (< $15) and at least allow you to do some basics like cook rice/pasta/soup/sauces.*
>
> *Finished the 13-week beginner program on DDP Yoga. Hit every workout on the schedule, didn't lag behind the schedule by even a single day once...Feels great to have stuck with it so well.*
>
> *I got flowers at work! And a sweet little card that said "I'm proud of you" from my mom because of how much effort and progress I've made in the last two months.*

Table 4 shows word semantic clusters that were significantly associated with lower weight loss, which are summarized below.

Table 4. Word semantic clusters that were statistically significant with respect to weight loss. Sample words are presented based on their distances to the center of a cluster in ascending order.

ID	Sample words	Beta	z score	SD	*P* value
C153	rising, dwindle, skyrocket, soar, rise, skyrockets, fluctuating, rises, fluctuates, spiking, spike, fluctuate	−.148	−2.782	0.053	.005
C359	purple, blue, colored, pink, brown, orange, burgundy, red, teal, gray, yellow, russet, tangerine, shades, amber	−.125	−3.464	0.036	.001
C120	request, permission, allotted, permit, requesting, requested, requests, alloted, permits, allotment, orders	−.114	−2.283	0.05	.02
C266	payments, payment, allowances, reimbursement, deduct, refund, deduction, benefits, premiums, payout	−.112	−3.282	0.034	.001
C184	vacation, vacations, getaway, vacay, resort, vacationing, resorts, honeymoon, hassle, hotels, hotel, disneyworld	−.108	−2.821	0.038	.005
C262	measurements, measurement, measuring, correlate, tracking, analysis, quantifying, monitoring, mapping	−.108	−3.627	0.030	<.001
C80	downstairs, sofa, upstairs, hallway, doorway, drawers, room, couch, cupboards, stairwell, bookshelf, fireplace	−.107	−2.022	0.053	.04
C418	hotline, contacting, contact, dialing, dialed, emailing, dial, calling, information, calls, info, referrals, referral	−.105	−5.653	0.019	<.001
C263	convince, persuade, entice, woo, coax, appease, tempt, lure, excite, inspire, convey, enlighten, quell, quench	−.101	−4.830	0.021	<.001
C421	production, producers, manufacturers, producing, output, prices, price, produce, pricing, manufacturer	−.091	−2.618	0.035	.009
C375	skinny, chubby, endomorph, lithe, pudgy, muscly, muscley, beefy, chunky, muscular, scrawny, lanky	−.084	−2.552	0.033	.01
C395	alternative, viable, workable, safe, safer, safest, alternatives, foolproof, alternate, bulletproof, inclusive	−.083	−3.455	0.024	.001
C103	tools, design, recommendations, recommended, designing, advice, recommend, materials, designs, kits	−.082	−3.075	0.027	.002
C20	split, dividing, separate, separating, splitting, divides, separated, divided, splits, divide, seperate, separation	−.070	−3.867	0.018	<.001
C227	league, scouts, rookie, leagues, teams, club, clubs, scout, seasons, season, veteran, rounder, draft	−.069	−2.911	0.024	.004
C142	budge, averse, succeeded, budging, budged, backslid, waver, incapable, qualms, bother, regressed, headway	−.062	−2.756	0.023	.006
C51	asda, morrisons, waitrose, sainsbury, sainsburys, weightwatchers, lidl, quorn, crisps, weetabix, twinings	−.060	−2.642	0.023	.008
C363	tonsils, cavities, tooth, pimples, molar, redness, toenails, teeth, sinuses, jaw, bone, zits, sinus, pimple	−.058	−2.232	0.026	.03
C18	varying, variety, varied, various, differing, different, multiple, individual, diverse, longevity, effectiveness	−.056	−3.412	0.016	.001
C124	undergraduate, college, undergrad, graduate, university, semester, graduates, semesters, academic, colleges	−.055	−2.117	0.026	.03
C92	maintenance, maintanence, maintence, maintenace, repairs, maintenence, maintainance, maintainence	−.050	−2.207	0.023	.03
C229	tweak, tweaking, tweaked, adjust, revising, updating, modifying, tweaks, modify, adjusting, readjust	−.046	−2.228	0.021	.03
C17	employed, working, work, volunteer, volunteering, volunteered, engaged, worked, engage, hire, hired	−.041	−2.127	0.019	.03
C333	moisturizer, lotion, creams, moisturizers, moisturizing, lotions, shampoo, cleanser, exfoliating, moisturising	−.034	−2.228	0.015	.03
C239	mother, daughter, aunt, niece, son, dad, grandmother, father, cousin, uncle, husband, wife, brother, sister	−.031	−2.133	0.014	.03

Activity-Related Clusters

Activity-related clusters included C184 (vacation), C227 (clubs, scouts), C124 (college, graduation), C92 (maintenance), and C17 (employment, volunteer). The following are some examples that were posted by users with lower weight loss:

> *I was fortunate enough to go on 2 European vacations this year (Italy in April, France and London 2 weeks ago) and I've gained 10lbs on top of the 15 I was already planning on losing.*
>
> *At my thinnest after college I was 5'6“ 165 lbs, and after I put 20 or 30 lbs back on my mom said she was relieved because I was starting to look ”too thin.“*
>
> *I'm actively seeking employment, so it depends what time I wake up if I have breakfast or not, but my usual day consists of...*

> *My weekend was okay. Had a big unhealthy meal yesterday but stayed 200 cal. under maintenance.*

Verb and Adjective Clusters

Verb and adjective clusters included C153 (skyrocket), C120 (request), C418 (contact hotline), C263 (entice related words), C20 (separate), C142 (backtrack), C229 (readjust), C359 (color), C375 (chunky), and C18 (various). The following are some examples:

> *I'm already exercising an hour 6 days a week. I really can't push it any more than that because my appetite skyrockets.*
>
> *I have three diagnosed illnesses and have been hospitalized twice. Users who exhibit suicidal behavior should be pointed to suicide prevention hotlines...*
>
> *I've backtracked by a couple weeks, which is partially water weight, and partially actual weight gain. It sucks.*
>
> *However, I'm over my 1200 for the day, not by much but I made a silly calorie budgeting decision earlier in the day. I readjusted dinner to try to make up for it, but I was too far in the hole.*
>
> *I was a chubby child, a chubby adult, and I've NEVER been in an acceptable BMI zone.*

Expense-Related Clusters

Expense-related clusters included C226 (payment, refund), C80 (home design), C51 (supermarket), C363 (otolaryngology related issues), C421 (production, price), C333 (moisturizer), and C239 (family members). The following are some examples that were related to these clusters:

> *I haven't been to the gym in over a week, and honestly don't think I will unless I can come up with $350 for the remainder of my Personal Training payments...oh, and more money for actual membership dues.*
>
> *I have enough motivation to get myself to the gym and am actually starting to have fun, but today I went after 4 weeks of not going, because I had my tonsils removed.*
>
> *Guilty of this! I tend to drink more frequently than my husband. I was going to try to stop drinking on weekdays. However last week, for whatever reason, he had a beer 4/7 nights.*

Discussion

Principal Findings

We used topic modeling to identify 25 general topics from the *r/loseit* subreddit. These topics covered a broad range of weight loss–related themes, including food and drinks, exercises, calorie, health issues, family members and friends, feelings, and communication. Among these topics, topics regarding food and drinks, health issues, family members and friends, calorie, and exercise were most discussed. These topics were aligned with the findings from another study [23].

Our regression analysis showed that the start weight and active days were associated with higher weight loss, which was aligned with our common sense. Furthermore, our results showed that receiving a higher karma score was associated with higher weight loss, but the median number of comments received was not significantly associated with weight loss. Our findings were a little different from the two studies, where both the karma score and comments were associated with higher weight loss [24,25]. We suspected that this might be because (1) we included far more users in our study and most of the posts received a very limited number of comments (Table 1) and (2) the previous studies did not control the model with more detailed content.

After adjusting for active days, start weight, karma score, and the median number of comments received, our analysis suggested that exercises, including coaching and nutrition, were the most effective content factors that were associated with higher weight loss, which were confirmed by previous investigations [33,34]. In addition, users with higher weight loss mentioned negative emotions that they experienced before they started to make efforts for weight loss. Our findings also suggested that mentioning food-related topics (eg, not eating too much, eating healthy food) were associated with higher weight loss, which was also found in a previous study [35]. Interestingly, we found that the mention of Xbox games was associated with higher weight loss as well. Evidence suggested that incorporating active video games had a positive effect on increasing physical activity and promoting healthy weight for both overweight adults and children [36,37].

In addition, we found that many content factors were associated with lower weight loss. For example, we found that people who mentioned vacations and clubs were more likely to have lower weight loss [38]. Furthermore, users in this subreddit mentioned that they gained weight after college graduation. Those users who had lower weight loss often mentioned supermarkets, payment or refund to exercise programs, and employment [39,40]. We also found that users who experienced health issues related to otolaryngology tended to have less weight loss. This might be due to the fact that the related treatment disturbed the weight loss plan. However, a study found that otorhinolaryngologic diseases themselves were associated with patients with obesity [41].

Another interesting finding was that users who used *skyrocket* to describe their weight loss experience (eg, feeling of eating) were less likely to have significant weight loss. This suggested that controlling the diet extensively during this process might not be an effective, healthy strategy [42]. In addition, users who often mentioned maintenance (eg, maintaining the intake of calories) were less likely to lose weight as expected. After a close examination of the related posts, we found that some of these users struggled with weight loss activities. We also found that expense-related content was associated with lower weight loss. This could be explained by a recent finding that low socioeconomic status was associated with lower weight loss outcomes [43]. Finally, users who mentioned family members were found less likely to lose more weight, suggesting that family members may not always have a positive impact on weight loss as found by other studies [44].

It should be noted that after examining feature correlation, only 6 topic predictors were included in the regression model, suggesting that word semantic clusters can capture more detailed offline activities. It was interesting that we found a calorie-related topic (T7) associated with lower weight loss, which could be partially supported by a previous finding that reducing calorie intake alone may not help in weight loss [45].

Implications

In this study, we acknowledged that while some associations were statistically significant, the value of the coefficients was very low, indicating a weak correlation between predictors and the dependent variable. However, we did not directly interpret the predictor importance from the values of their coefficients. This is because it is practically meaningless to say that *more weight can be reduced by increasing the distribution of a certain topic discussed in an online community* [46]. Rather, we believe that it is the actual offline activities described in online discussions (or self-efficacy) that actually matter in weight loss. By using word clusters, we obtained more detailed, concrete offline activities that were often ignored by other social media–based studies but were significantly associated with the amount of weight loss. While karma scores (votes) from other users were associated with higher weight loss, considering the right-skewed distribution of karma scores (Table 1 and Multimedia Appendix 1), a majority of posts in this online community received very small karma scores.

These were somewhat aligned with the findings in an offline, SCT-based weight loss intervention program [47], where self-efficacy and intention, instead of online interaction, were found to be significant factors leading to weight loss. While self-efficacy performed well in weight loss interventions [48-50], there was evidence that self-efficacy may face the challenge of decreasing over time [51]. This is very interesting because it implied that interactions, either in online or offline environments, may serve as an indirect factor that affects weight loss through maintaining a participant’s long-term self-efficacy.

From this perspective, future weight loss analyses or interventions should consider online interaction as a key factor to improve self-efficacy, instead of directly being linked to weight loss. Our study also implied that an aggressive weight loss plan may not work in the long run.

Limitations and Future Work

There are several limitations that we want to highlight here. First, our findings were based on merely the *r/loseit* subreddit, which constrained the generality of findings. Future work may consider extending the research to other online platforms. Second, we did not incorporate gender into the analysis. It might be possible to first infer such information from online discussions [52] and then investigate how the association between posting content and weight loss changes after controlling for this information. Third, we relied on self-reported weight loss in this study, which limited our investigation, and findings were applicable to only a small fraction of Reddit users who disclosed weight change. It will be interesting to investigate the specific characteristics that are related to the majority of Reddit users who did not report such information. Furthermore, self-reported weight changes might not be accurate because Reddit users might not update their weights in a timely manner. Our study only investigated what was presented in the online discussions, instead of examining the real-world events. Finally, it would be interesting to investigate the extent to which online interaction, not merely responses and votes but the detailed categories, and offline activities that were recorded in online discussions can lead to weight loss change in a dynamic setting.

Conclusions

In this study, we analyzed online discussions regarding weight loss in *r/loseit*. We used topic modeling and the hierarchical clustering algorithm to extract topics and word clusters that were discussed in this subreddit. We used a regression analysis to determine the association between weight loss change and the factors that were conveyed in these online discussions. We found that the start weight and median karma scores were associated with higher weight loss. Users who had higher weight loss might be motivated by negative emotions experienced before starting weight loss. By contrast, users who mentioned vacations and payments were less likely to lose more weight. Furthermore, mentions of family members and employment were also found to be associated with lower weight loss. Our findings suggest that future interventions based on online social media platforms should focus on both online interaction and offline activities and that online personal health data can be effectively used to learn about users’ health-related behaviors.

Acknowledgments

The authors thank the colleagues in Changchun Normal University and Vanderbilt University for the positive feedback on this project. Special thanks go to Na Liu from Georgia Institute of Technology for the constructive discussion on the work implication.

Authors' Contributions

ZY proposed the research question and designed the experiments. YL collected data and conducted the experiments. ZY and YL wrote the manuscript. ZY reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental materials for selection of the optimal number of topics and the distribution of karma score.
[DOCX File , 713 KB - jmir_v22i6e13745_app1.docx]

References

1. Buchwald H, Estok R, Fahrbach K, Banel D, Jensen MD, Pories WJ, et al. Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. Am J Med 2009 Mar;122(3):248-56.e5. [doi: 10.1016/j.amjmed.2008.09.041] [Medline: 19272486]
2. Parker ED, Sinaiko AR, Kharbanda EO, Margolis KL, Daley MF, Trower NK, et al. Change in weight status and development of hypertension. Pediatrics 2016 Mar;137(3):e20151662 [FREE Full text] [doi: 10.1542/peds.2015-1662] [Medline: 26908707]
3. Chlebowski RT, Luo J, Anderson GL, Barrington W, Reding K, Simon MS, et al. Weight loss and breast cancer incidence in postmenopausal women. Cancer 2019 Jan 15;125(2):205-212 [FREE Full text] [doi: 10.1002/cncr.31687] [Medline: 30294816]
4. Kovesdy CP, Furth SL, Zoccali C, World Kidney Day Steering Committee. Obesity and kidney disease: hidden consequences of the epidemic. Nephron 2017;135(4):243-251. [doi: 10.1159/000455698] [Medline: 28171864]
5. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. N Engl J Med 2000 Jul 6;343(1):16-22. [doi: 10.1056/NEJM200007063430103] [Medline: 10882764]
6. Chiuve SE, Rexrode KM, Spiegelman D, Logroscino G, Manson JE, Rimm EB. Primary prevention of stroke by healthy lifestyle. Circulation 2008 Aug 26;118(9):947-954 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.781062] [Medline: 18697819]
7. Centers for Disease Control and Prevention. 2019. Adult Obesity Facts URL: https://www.cdc.gov/obesity/data/adult.html [accessed 2019-12-21]
8. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A scalable framework to detect personal health mentions on Twitter. J Med Internet Res 2015 Jun 5;17(6):e138 [FREE Full text] [doi: 10.2196/jmir.4305] [Medline: 26048075]
9. Mejova Y, Haddadi H, Noulas A. #FoodPorn: Obesity Patterns in Culinary Interactions. In: Proceedings of the 5th International Conference on Digital Health. 2015 Presented at: DH'15; May 18 - 20, 2015; Florence, Italy p. 51-58. [doi: 10.1145/2750511.2750524]
10. Cesare N, Dwivedi P, Nguyen QC, Nsoesie EO. Use of social media, search queries, and demographic data to assess obesity prevalence in the United States. Palgrave Commun 2019 Sep 17;5(106):1-9. [doi: 10.1057/s41599-019-0314-x]
11. Turner-McGrievy GM, Beets MW. Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. Transl Behav Med 2015 Jun;5(2):160-166 [FREE Full text] [doi: 10.1007/s13142-015-0308-1] [Medline: 26029278]
12. Pagoto S, Schneider KL, Evans M, Waring ME, Appelhans B, Busch AM, et al. Tweeting it off: characteristics of adults who tweet about a weight loss attempt. J Am Med Inform Assoc 2014;21(6):1032-1037 [FREE Full text] [doi: 10.1136/amiajnl-2014-002652] [Medline: 24928175]
13. Reading JM, Buhr KJ, Stuckey HL. Social experiences of adults using online support forums to lose weight: a qualitative content analysis. Health Educ Behav 2019 Dec;46(2_suppl):129-133. [doi: 10.1177/1090198119859403] [Medline: 31742458]
14. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31 [FREE Full text] [doi: 10.1002/oby.20232] [Medline: 23505165]
15. Merchant G, Weibel N, Patrick K, Fowler JH, Norman GJ, Gupta A, et al. Click 'like' to change your behavior: a mixed methods study of college students' exposure to and engagement with Facebook content designed for weight loss. J Med Internet Res 2014 Jun 24;16(6):e158 [FREE Full text] [doi: 10.2196/jmir.3267] [Medline: 24964294]
16. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
17. Hales SB, Davidson C, Turner-McGrievy GM. Varying social media post types differentially impacts engagement in a behavioral weight loss intervention. Transl Behav Med 2014 Dec;4(4):355-362 [FREE Full text] [doi: 10.1007/s13142-014-0274-z] [Medline: 25584084]
18. Maher CA, Lewis LK, Ferrar K, Marshall S, de Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
19. Hagger MS, Weed M. DEBATE: Do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019 Apr 25;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
20. Hwang KO, Ning J, Trickey AW, Sciamanna CN. Website usage and weight loss in a free commercial online weight loss program: retrospective cohort study. J Med Internet Res 2013 Jan 15;15(1):e11 [FREE Full text] [doi: 10.2196/jmir.2195] [Medline: 23322819]

21. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of internet website use with weight change in a long-term weight loss maintenance program. J Med Internet Res 2010 Jul 27;12(3):e29 [FREE Full text] [doi: 10.2196/jmir.1504] [Medline: 20663751]
22. Brindal E, Freyne J, Saunders I, Berkovsky S, Smith G, Noakes M. Features predicting weight loss in overweight or obese participants in a web-based intervention: randomized trial. J Med Internet Res 2012 Dec 12;14(6):e173 [FREE Full text] [doi: 10.2196/jmir.2156] [Medline: 23234759]
23. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Silva AP, Meira W, et al. Factors associated with weight change in online weight management communities: a case study in the loseit reddit community. J Med Internet Res 2017 Jan 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]
24. Cunha T, Weber I, Pappa G. A Warm Welcome Matters!: The Link Between Social Feedback and Weight Loss in /r/loseit. In: Proceedings of the 26th International Conference on World Wide Web Companion. 2017 Presented at: WWW'17 Companion; April 3 - 7, 2017; Perth, Australia p. 1063-1072. [doi: 10.1145/3041021.3055131]
25. Cunha TO, Weber I, Haddadi H, Pappa GL. The Effect of Social Feedback in a Reddit Weight Loss Community. In: Proceedings of the 6th International Conference on Digital Health Conference. 2016 Presented at: DH'16; April 11 - 13, 2016; Québec, Montréal, Canada p. 99-103. [doi: 10.1145/2896338.2897732]
26. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]
27. Shamizadeh T, Jahangiry L, Sarbakhsh P, Ponnet K. Social cognitive theory-based intervention to promote physical activity among prediabetic rural people: a cluster randomized controlled trial. Trials 2019 Feb 4;20(1):98 [FREE Full text] [doi: 10.1186/s13063-019-3220-z] [Medline: 30717779]
28. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and Their Compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. 2013 Presented at: NIPS'13; December 5-10, 2013; Lake Tahoe p. 3111-3119 URL: http://dl.acm.org/citation.cfm?id=2999792.2999959
29. Yin Z, Harrell M, Warner JL, Chen Q, Fabbri D, Malin BA. The therapy is making me sick: how online portal communications between breast cancer patients and physicians indicate medication discontinuation. J Am Med Inform Assoc 2018 Nov 1;25(11):1444-1451. [doi: 10.1093/jamia/ocy118] [Medline: 30380083]
30. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
31. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring Topic Coherence Over Many Models and Many Topics. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012 Presented at: EMNLP-CoNLL'12; July 12-14, 2012; Jeju Island, Korea p. 952-961 URL: https://dl.acm.org/doi/10.5555/2390948.2391052
32. De Choudhury M, De S. Mental Health Discourse on Redditlf- Self Disclosure, Social Support, and Anonymity. In: Association for the Advancement of Artificial Intelligence!. 2014 May 16 Presented at: ICWSM'14; June 1–4, 2014; Ann Arbor, Michigan, USA p. 71-80 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8075
33. Foster-Schubert KE, Alfano CM, Duggan CR, Xiao L, Campbell KL, Kong A, et al. Effect of diet and exercise, alone or combined, on weight and body composition in overweight-to-obese postmenopausal women. Obesity (Silver Spring) 2012 Aug;20(8):1628-1638 [FREE Full text] [doi: 10.1038/oby.2011.76] [Medline: 21494229]
34. Painter SL, Ahmed R, Kushner RF, Hill JO, Lindquist R, Brunning S, et al. Expert coaching in weight loss: retrospective analysis. J Med Internet Res 2018 Mar 13;20(3):e92 [FREE Full text] [doi: 10.2196/jmir.9738] [Medline: 29535082]
35. Soeliman FA, Azadbakht L. Weight loss maintenance: A review on dietary related strategies. J Res Med Sci 2014 Mar;19(3):268-275 [FREE Full text] [Medline: 24949037]
36. Trost SG, Sundal D, Foster GD, Lent MR, Vojta D. Effects of a pediatric weight management program with and without active video games a randomized trial. JAMA Pediatr 2014 May;168(5):407-413. [doi: 10.1001/jamapediatrics.2013.3436] [Medline: 24589566]
37. Staiano AE, Abraham AA, Calvert SL. The Wii Club: Gaming for weight loss in overweight and obese youth. Games Health J 2012 Oct 5;1(5):377-380 [FREE Full text] [doi: 10.1089/g4h.2012.0052] [Medline: 23565349]
38. Díaz-Zavala RG, Castro-Cantú MF, Valencia ME, Álvarez-Hernández G, Haby MM, Esparza-Romero J. Effect of the holiday season on weight gain: a narrative review. J Obes 2017;2017:2085136 [FREE Full text] [doi: 10.1155/2017/2085136] [Medline: 28744374]
39. Mutsaerts MA, Kuchenbecker WK, Mol BW, Land JA, Hoek A. Dropout is a problem in lifestyle intervention programs for overweight and obese infertile women: a systematic review. Hum Reprod 2013 Apr;28(4):979-986. [doi: 10.1093/humrep/det026] [Medline: 23427235]
40. Lemstra M, Bird Y, Nwankwo C, Rogers M, Moraros J. Weight loss intervention adherence and factors promoting adherence: a meta-analysis. Patient Prefer Adherence 2016;10:1547-1559 [FREE Full text] [doi: 10.2147/PPA.S103649] [Medline: 27574404]
41. Kim TH, Kang HM, Oh I, Yeo SG. Relationship between otorhinolaryngologic diseases and obesity. Clin Exp Otorhinolaryngol 2015 Sep;8(3):194-197 [FREE Full text] [doi: 10.3342/ceo.2015.8.3.194] [Medline: 26330910]

42. Institute of Medicine, Food and Nutrition Board, Committee on Military Nutrition Research, Subcommittee on Military Weight Management. Weight Management: State of the Science and Opportunities for Military Programs. Washington, DC: National Academies Press; 2004.
43. Carden A, Blum K, Arbaugh CJ, Trickey A, Eisenberg D. Low socioeconomic status is associated with lower weight-loss outcomes 10-years after Roux-en-Y gastric bypass. Surg Endosc 2019 Feb;33(2):454-459. [doi: 10.1007/s00464-018-6318-6] [Medline: 29987570]
44. Wang ML, Pbert L, Lemon SC. Influence of family, friend and coworker social support and social undermining on weight gain prevention among adults. Obesity (Silver Spring) 2014 Sep;22(9):1973-1980 [FREE Full text] [doi: 10.1002/oby.20814] [Medline: 24942930]
45. Benton D, Young HA. Reducing calorie intake may not help you lose body weight. Perspect Psychol Sci 2017 Sep;12(5):703-714 [FREE Full text] [doi: 10.1177/1745691617690878] [Medline: 28657838]
46. Yin Z, Sulieman L, Malin B. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc 2019 Jun 1;26(6):561-576. [doi: 10.1093/jamia/ocz009] [Medline: 30908576]
47. Young MD, Plotnikoff RC, Collins CE, Callister R, Morgan PJ. A test of social cognitive theory to explain men's physical activity during a gender-tailored weight loss program. Am J Mens Health 2016 Nov;10(6):NP176-NP187. [doi: 10.1177/1557988315600063] [Medline: 26275722]
48. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? an in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
49. Brawley L, Rejeski WJ, Gaukstern JE, Ambrosius WT. Social cognitive changes following weight loss and physical activity interventions in obese, older adults in poor cardiovascular health. Ann Behav Med 2012 Dec;44(3):353-364 [FREE Full text] [doi: 10.1007/s12160-012-9390-5] [Medline: 22773225]
50. Nezami BT, Lang W, Jakicic JM, Davis KK, Polzien K, Rickman AD, et al. The effect of self-efficacy on behavior and weight in a behavioral weight-loss intervention. Health Psychol 2016 May 16:- [Epub ahead of print] [FREE Full text] [doi: 10.1037/hea0000378] [Medline: 27183306]
51. McAuley E, Mailey EL, Mullen SP, Szabo AN, Wójcicki TR, White SM, et al. Growth trajectories of exercise self-efficacy in older adults: influence of measures and initial status. Health Psychol 2011 Jan;30(1):75-83 [FREE Full text] [doi: 10.1037/a0021567] [Medline: 21038962]
52. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS One 2013;8(9):e73791 [FREE Full text] [doi: 10.1371/journal.pone.0073791] [Medline: 24086296]

Abbreviations

LDA: latent Dirichlet allocation
SCT: social cognitive theory

Edited by G Eysenbach; submitted 15.03.19; peer-reviewed by S Chancellor, G Petrič, S Atanasova; comments to author 24.09.19; revised version received 24.12.19; accepted 26.02.20; published 08.06.20.

Please cite as:
Liu Y, Yin Z
Understanding Weight Loss via Online Discussions: Content Analysis of Reddit Posts Using Topic Modeling and Word Clustering Techniques
J Med Internet Res 2020;22(6):e13745
URL: https://www.jmir.org/2020/6/e13745
doi:10.2196/13745
PMID:32510460

Original Paper

Superusers’ Engagement in Asthma Online Communities: Asynchronous Web-Based Interview Study

Anna De Simoni[1], PhD; Anjali T Shah[2], BChir, MB; Olivia Fulton[3], BSc; Jasmine Parkinson[4], DPhil; Aziz Sheikh[5], MD; Pietro Panzarasa[6], PhD; Claudia Pagliari[7], PhD; Neil S Coulson[8], PhD; Chris J Griffiths[1], DPhil

[1]Asthma UK Centre for Applied Research, Institute of Population Health Sciences, Queen Mary University of London, London, United Kingdom
[2]University of Cambridge, Cambridge, United Kingdom
[3]Asthma UK Centre for Applied Research, University of Edinburgh, Edinburgh, United Kingdom
[4]Asthma UK, London, United Kingdom
[5]Asthma UK Centre for Applied Research, Usher Institute of Population Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom
[6]School of Business and Management, Queen Mary University of London, London, United Kingdom
[7]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom
[8]School of Medicine, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Anna De Simoni, PhD
Asthma UK Centre for Applied Research
Institute of Population Health Sciences
Queen Mary University of London
58 Turner street
London,
United Kingdom
Phone: 44 207 882 2520
Email: a.desimoni@qmul.ac.uk

Abstract

Background: Superusers, defined as the 1% of users who write a large number of posts, play critical roles in online health communities (OHCs), catalyzing engagement and influencing other users’ self-care. Their unique online behavior is key to sustaining activity in OHCs and making them flourish. Our previous work showed the presence of 20 to 30 superusers active on a weekly basis among 3345 users in the nationwide Asthma UK OHC and that the community would disintegrate if superusers were removed. Recruiting these highly skilled individuals for research purposes can be challenging, and little is known about superusers.

Objective: This study aimed to explore superusers’ motivation to actively engage in OHCs, the difficulties they may face, and their interactions with health care professionals (HCPs).

Methods: An asynchronous web-based structured interview study was conducted. Superusers of the Asthma UK OHC and Facebook groups were recruited through Asthma UK staff to pilot and subsequently complete the questionnaire. Open-ended questions were analyzed using content analysis.

Results: There were 17 superusers recruited for the study (14 patients with asthma and 3 carers); the majority were female (15/17). The age range of participants was 18 to 75 years. They were active in OHCs for 1 to 6 years and spent between 1 and 20 hours per week reading and 1 and 3 hours per week writing posts. Superusers’ participation in OHCs was prompted by curiosity about asthma and its medical treatment and by the availability of spare time when they were off work due to asthma exacerbations or retired. Their engagement increased over time as participants furthered their familiarity with the OHCs and their knowledge of asthma and its self-management. Financial or social recognition of the superuser role was not important; their reward came from helping and interacting with others. According to the replies provided, they showed careful judgment to distinguish what can be dealt with through peer advice and what needs input from HCPs. Difficulties were encountered when dealing with misunderstandings about asthma and its treatment, patients not seeking advice from HCPs when needed, and *miracle cures* or dangerous ideas. Out of 17 participants, only 3 stated that their HCPs were aware of their engagement with OHCs. All superusers thought that HCPs should direct patients to OHCs, provided they are trusted and moderated. In addition, 9 users felt that HCPs themselves should take part in OHCs.

Conclusions: Superusers from a UK-wide online community are highly motivated, altruistic, and mostly female individuals who exhibit judgment about the complexity of coping with asthma and the limits of their advice. Engagement with OHCs satisfies their psychosocial needs. Future research should explore how to address their unmet needs, their interactions with HCPs, and the potential integration of OHCs in traditional healthcare.

(J Med Internet Res 2020;22(6):e18185) doi:10.2196/18185

KEYWORDS

social networks; eHealth; social media; peer-to-peer support; social support; online health communities; online forums; superusers; leadership; misinformation; asthma; self-management

Introduction

Background

Recent work has suggested that taking part in online communities for people with long-term conditions (LTCs) improves illness self-management [1] and adherence to treatment [2], produces positive health-related outcomes [3-5], facilitates shared decision making with health care professionals (HCPs) [6-8], and may even reduce mortality [9]. There is also evidence that self-management support interventions can reduce health service utilization [10,11]. Participation in online health communities (OHCs) for patients with LTCs can take up part of the health care service demand and indirectly improve access to health care [12]. However, much of this evidence comes from qualitative and observational studies [6,13,14]. Despite a lack of definitive evidence, policymakers are starting to see the potential of OHCs, for example, the Big White Wall, an OHC commissioned by some mental health services in the United Kingdom, Canada, and New Zealand [15]. The Irish health system is piloting closed Facebook groups for smoking cessation [16], whereas the Public Health England Stoptober smoking cessation campaign includes a Facebook group [17] among other initiatives. In the United Kingdom, Facebook is also being piloted by the National Health System Digital (NHS Digital) to improve cancer screening rates, with promising results [18]. This increasing attention to health social media calls for elucidating the mechanisms that make OHC engagement successful in terms of improving self-management [12,19]. Indeed, although some OHCs flourish, many suffer from little or no traffic [20]. The emerging literature investigating mechanisms of effective OHC engagement shows that superusers (ie, users who are in the top 1%-5% in terms of messages posted to the OHC) are key to success. They generate the majority of traffic and create value, so their recruitment and retention are imperative for the long-term success of OHCs [21]. A previous analysis of an online community for people with drinking problems found that common themes for superusers' engagement included introductions, greetings, general supportive statements, suggested strategies, success stories, and discussion of difficulties [22], showing that superusers reassuringly offer peer support toward behavioral and emotional self-management tasks [23], appropriately leaving to HCPs medical self-management tasks.

To fully understand the unique mechanisms of behavior change through internet-based interventions, collaboration and knowledge transfer between researchers, nonprofit organizations, and private organizations have been recommended [24]. Our network study of peer support in the Asthma UK and British Lung Foundation (BLF) OHCs [25] in collaboration with the platform provider HealthUnlocked has highlighted the key role of superusers. Superusers are distinct from community moderators formally appointed by the platform provider; the number of moderators among the highly active users was negligible [25]. Superusers are a naturally available resource and are responsible for holding successful OHCs together, engage with users with low posting activity, and indirectly contribute to the formation of ties between users. As users become more active within the community, they become more likely to reply to posts than to ask questions. This suggests that superusers gradually become *experts*, providing others with advice and support [2,7,26].

Online superusers could be considered allies of the health care workforce [27,28]. Our work has inspired the development of a new network-based theory of social medical capital, broadly defined in terms of the advantages that any user (patient or caregiver) can gain from participation in OHCs [29].

In this context, strategies to increase superusers' participation can improve engagement with OHCs [30]. This is attracting growing interest from academics, HCPs, and policymakers. Despite the evidence of superusers' key role in successful OHCs, little is known about this small but critical population, what motivates them to contribute to the community and stay active over time [22,25], whether they encounter any challenges, how their contribution could be supported in any way, and what would make OHCs safer and more effective. This study set out to understand what motivates superusers to adopt this role and the reciprocal value it offers them. Although the research was pragmatically driven, our interpretation draws on self-determination theory [31], a framework for differentiating intrinsic and extrinsic forms of motivation, to help link the insights to recommendations for health organizations seeking to engage with and utilize the value offered by superusers in their online communities.

Objectives

Here, we undertook an asynchronous web-based structured interview study of UK superusers previously quantitatively characterized [25], in collaboration with the charity Asthma UK. We additionally aimed to explore superusers' interfaces with HCPs and their views on HCPs' potential role in promoting engagement with OHCs and on HCPs' engagement as OHC participants themselves.

Methods

Interview Schedule Development

The interview schedule was developed with questions based on the extant literature, recent work on OHCs [25,32], and informal discussion dated from 2015 to 2018 with 2 superusers, 1 from a stroke OHC previously studied [33] and 1 from Asthma UK OHC.

Piloting Phase

Piloting was undertaken with 6 superusers recruited through the Asthma UK research operating officer (JP) and OHC moderator. JP emailed the weblink to the study questions and attached to the same email a Microsoft Word document with the interview questions in November 2018. Comments and suggestions were received by JP between November 2018 and February 2019. Superusers' suggestions improved the clarity of the introductory text and the queries asked. Some questions initially part of the same query were split to make replying easier (ie, questions 3-5 and 7-8), whereas new questions were suggested (ie, questions 10, 13-15, 21, and 24). This process resulted in 10 additional questions. The wording of some questions was also adjusted to make it more neutral to participants.

Inclusion Criteria

The inclusion criteria were as follows:

- Living with asthma or caring for somebody with asthma
- Having posted to an online asthma community at least one message per week for at least four weeks.

As there is no evidence yet about whether superusers' posting activity over time is regular or occurring in bursts (eg, when off work due to illness), we opted to be nonspecific about the 4-week period. Therefore, posting activity over any 4 weeks, consecutive or not, at any point in time would qualify participants as superusers.

In this asynchronous web-based structured interview, the definition of superusers is different from the *retrospective* one used in the study by Joglekar et al [25] (ie, top 1% of users characterized by the largest number of posts posted in the community *over the entire observation period of 10 years).* This previous study showed that only about 20 to 30 superusers were active on a weekly basis since 2015. The inclusion criteria for superusers were agreed upon by coauthors and the superusers who took part in the pilot phase. Due to the hypothesized small sample of potential participants available to recruitment [25], no saturation criteria were used to determine the study sample size.

Participant Recruitment

Of the 17 participants, 16 were recruited by the Asthma UK research operating officer (JP) by email and through an Asthma UK monthly email bulletin to take part anonymously through a SurveyMonkey link [34]. Responses were collected between March and April 2019. A superuser (OF) who is a member of the Asthma UK Centre for Applied Research Patient and Public Involvement (PPI) group [35] was invited and recruited by AD.

Ethical Approval

The study was approved by the Queen Mary University Research Ethics Committee (ref QMREC2205a). To address the issue of confidentiality around patient information and to avoid this information being known to the research team, superusers were approached only by the Asthma UK staff (JP) and invited to participate. The research team did not have access to personally identifiable information apart from the AUKAR PPI member and coauthor (OF).

Analysis

We analyzed the text from open questions using inductive content analysis as described by Elo and Kyngas [36]. Two authors (AS and AD) read all responses to familiarize themselves with the data. An initial coding framework with themes and subthemes was developed, which was adjusted as new data were added. This was done for the first 10 individuals and, subsequently, for the additional 7 individuals. Coding was then performed independently by 2 authors (AS and AD) on all data. Coding was discussed until agreement was reached, and the themes were revised as well.

Results

Characteristics of the Participants

A total of 17 participants were included in the study (Table 1): 14 were people living with asthma, whereas 3 were mothers of children with asthma. The majority were female (15/17), with an age range of 18 to 75 years (3 out of 17 participants were aged 66-75 years).

Of the 17 users, 10 participated in 2 or more OHCs: 15 out of 17 in Asthma UK HealthUnlocked community and 10 out of 15 in Facebook groups. HealthUnlocked is the platform provider of the Asthma UK online community.

With respect to education, 65% (11/17) had at least an undergraduate degree and 18% (3/17) had a postgraduate degree.

Before taking part in the study, they had been active in OHCs for 1 to 6 years and spent between 1 and 20 hours/week (11 out 17 participants spent ≥2 hour/week) reading and between 1 and 3 hours/week (7 out of 17 participants spent ≥1 hour/week) writing posts.

Self-reported participation increased over time for 14 out of 17 superusers and was linked to wanting to know more about asthma and its treatment in the context of deterioration of asthma or change in medical treatment. Other factors contributing to participation were increased familiarity and interest toward OHC members and improved awareness and knowledge of asthma.

Table 1. Participants' characteristics.

Participant number	Gender	Age range (years)	OHCs[a] joined	Duration as OHC member (years)	Time spent reading posts[b] (hours)	Time spent writing posts[b] (hours)	Number of posts written[b]	Highest level of education
1[c]	Female	36-45	Facebook group (UK Parents of Children with Asthma)	2	5	<1	4-5	Undergraduate degree or similar
2	Female	46-55	Asthma UK	1	1	N/A[d]	0	Postgraduate degree or similar (eg, PhD)
3	Female	66-75	Asthma UK	>1	5	NS[e]	4-10	Undergraduate degree or similar
4	Female	66-75	Asthma UK	2.5	>0.5[f]	NS[e]	Variable[f]	Postgraduate degree or similar (eg, PhD)
5	Female	66-75	British Lung Foundation, Asthma UK	>3	10	2	4	Undergraduate degree or similar
6	Female	46-55	Asthma UK, British Lung Foundation	5	5	<0.5	1	O-levels/ General Certificate of Secondary Education or similar
7	Female	46-55	Asthma UK	5	20	1	5	Prefer not to say
8	Male	46-55	Asthma UK, Facebook	6	1	NS	1	A-level or similar
9	Male	46-55	Asthma UK	0.5	1	0.3	2	Undergraduate degree or similar
10	Female	18-25	Facebook groups, Asthma UK	0.5	1-2	0	0-1	Undergraduate degree or similar
11	Female	46-55	Asthma UK	>3	1-2	1	0-1	A-level or similar
12[c]	Female	36-45	Asthma UK, Facebook group (UK Parents of Children with Asthma)	2.5	2	0.3	4	Undergraduate degree or similar
13	Female	26-35	Asthma UK, HealthUnlocked communities	2.5	6	2-3	15	Postgraduate degree or similar (eg, PhD)
14	N/A	N/A	Facebook group (Bronchiectasis)	<1	0.5	NS	0-5	NS
15	Female	26-35	Asthma UK, Facebook groups	>5	3	1-2	3	Undergraduate degree or similar
16[c]	Female	36-45	Facebook group (UK Parents of Children with Asthma), Asthma UK	2-3	2	0.2	5	Undergraduate degree or similar
17	Female	18-25	Asthma UK, Facebook groups	0.5	3-4	0-1	0-1	A-level or similar

[a] OHCs: online health communities.

[b] Hours of engagement and number of posts refer here to the average week. Some superusers' engagement may be concentrated over certain time periods.

[c] Superuser is a carer (mother of a child with asthma).

[d] N/A: not applicable.

[e] NS: not stated.

[f] Higher contribution when not well with asthma.

Themes

Themes and subthemes were generated through content analysis of open-ended questions and are shown in Table 2. Our findings will be articulated into 4 themes:

1. Motivation to engage: Motivation to active participation in OHCs included personal advantage and the desire to help others/being altruistic. Engagement with OHC promoted superusers' sense of personal control, agency (ie, the actual ability to deal with a task or situation), and self-efficacy (ie, the perceived ability to deal with a task or situation) over their illness, particularly when they adopted the informal role of *wise mentors* wise mentors or supporters to other users. An important reason for people taking part in asthma OHCs was the reward felt by being helpful to other members.
2. Awareness of the limits of peer self-management support: Superusers showed a sense of awareness of what can be dealt with through peer support/advice (ie, behavioral and emotional self-management tasks) and of what instead would require input from HCPs (ie, medical self-management tasks).
3. Challenges: Superusers encounter several challenges, such as dealing with misunderstanding about asthma and treatment and a general sense of responsibility toward other OHC users. Superusers could find themselves under moral pressure to respond to risk-indicating or inappropriate posts and reporting posts to moderators, which could cause at times cognitive stress.
4. Interface with HCPs: Most HCPs were not aware of OHC engagement. Superusers generally felt that HCPs should direct patients to well-trusted and moderated OHCs. Some felt that HCPs themselves should take part in OHCs.

Table 2. Themes.

Themes	Subthemes
Motivation	
Seeking information and support	• Gaining knowledge about asthma illness and its treatment • Validation of own experiences in the context of asthma • Feeling less isolated • Talking with fellow sufferers • Enjoying reading conversations of users one got to know
Helping others	• Giving advice that could potentially save a life • Supporting others with asthma and their carers • Making people with asthma and their carers feeling less isolated • Show others that living with chronic conditions is not always negative • Making sure people with asthma take their disease seriously • Disseminating of scientific information
Feeling rewarded through helping	• Positive feeling when helping others • Replacement of role
Financial or social recognition not important	• Not important: reward is helping others, enjoying interactions • Not important: all users shall be equal with equal voice importance • Not important: voluntary nature—financial incentives potentially causing misuse • Unsure/yes important: recognition of sensible/knowledgeable users
Awareness of the limits of peer self-management support	
Decisions on posts to reply to	• Drawing from personal experience • Providing a different point of view
Types of support offered	• Behavioral and emotional peer self-management support • Empowering patients and carers through own experience • Signposting to source information and support
Medical self-management needs HCPs’[a] input	• Showing appropriate insight of potentially serious medical issues • Advising to seek medical help with appropriate urgency
Challenges	
Top 3 problems and difficulties encountered	• Worrying about users struggling with asthma and not seeking medical help as appropriate • Misunderstanding, spam, miracle cures, or dangerous ideas • Negative tone of some conversations
Posts causing superusers’ worries	• Religion-based advocacies; derogatory, emotionally challenging posts • Offering bad advice or indicating that users have little knowledge about asthma and its gravity
Need of additional policies and guidance	• Improve awareness of existing policies and guidance for safe engagement with asthma OHCs[b] • Policies and guidance about buying asthma medications on the web • Quicker removal of bad posts/advice
HCPs and asthma OHCs	
HCPs’ awareness of engagement with asthma OHCs	• Engagement with asthma OHCs is not discussed during consultations with HCPs • Engagement with asthma OHCs is not discouraged by HCPs • HCPs’ belief that engagement with asthma OHCs focuses patients on illness and potentially increases their anxiety

Themes	Subthemes
HCPs' promotion of engagement with OHCs	HCPs should direct patients with LTCs[c] to moderated/trusted OHCs Advantage is obtaining behavioral and emotional self-management support that HCPs may not be able to offer Ways of promotion: posters up in the waiting rooms of relevant hospital departments and General Practice (GP) surgeries and face-to-face discussion of OHC information by nurses during asthma clinics
Suggestions to reassure HCPs about OHC engagement	Clearer statements about contacting HCPs for medical self-management Improving HCPs' awareness of benefits of online peer support; robust evaluation of the effects of OHC engagement Readily accessible guidance about keeping safe in social media Clear rules about posting activity; regular, nonintrusive participation of moderators
HCPs' participation in OHCs	Benefits: for the opportunity to get worries and questions addressed, as long as HCPs' identity is stated Difficulties: potential scrutiny of all posts, limitation of expression of different points of view, lack of sufficient clinical details, and issues with code of conduct of HCP registering bodies

[a] HCP: health care professional.

[b] OHC: online health community.

[c] LTC: long-term condition.

Motivation of Engagement

Seeking Information/Support

Motivation to engage with OHCs was linked to personal advantage through gaining knowledge and support for asthma and its treatment:

> *To learn from others who actually know what it can be like and to learn from their experiences.* [N.4]

Validation of own experiences in the context of asthma and the feeling of being less isolated were also important factors:

> *To get validation from others with the same symptoms.* [N.17]

Having the opportunity to talk with people who live with asthma was considered important:

> *...many people find comfort and support in such communities that cannot be offered by family members and/or friends that have not experienced the day-to-day living of conditions.* [N.14]

Reading other users' conversation was described as a positive experience that increased engagement:

> *I enjoy the chats with others and reading the dialogue between others, many of whom I've got to know.* [N.5]

Helping Others

Altruisms and the benefit of feeling in a position to help others were a significant factor sustaining the motivation to regularly take part in OHCs.

Some even mentioned the potential to save lives:

> *If my own message or experience could help save a child's life.* [N.1]

Most participants talked about the motivation to support others with asthma and their carers to mitigate their sense of social isolation:

> *…especially when you are supporting parents who are new to dealing with asthma in their child.* [N.1]
>
> *I remember how lost and frustrated and alone I felt. I don't want other mums who are new to asthma to have those same feelings and struggles.* [N.16]
>
> *...being there to give emotional support when needed.* [N.17]

Part of the motivation was to show others that living with chronic conditions is not always negative and to offer hope to others:

> *I wish to show that I can live a normal life with chronic condition.* [N.2]

Using their knowledge to clear up any confusion about asthma and medications was relevant, as well as making sure people with asthma took their disease seriously and did not rely on social media for queries that needed HCPs' input:

> *Trying to make people take their asthma more seriously and not rely on social media for the answers which often don't come and then they end up in hospital.* [N.15]

Interestingly, a participant mentioned that part of the motivation was to disseminate proper scientific information:

> *To be helpful and disseminate information, especially scientific information.* [N.8]

Rewards for Online Health Communities Engagement

Participants found helping others a positive experience for themselves. By providing replies to other users' queries, superusers increasingly acquired confidence and were recognized for their role as community *experts*, which in turn boosted their motivation to further engage in OHCs. For some participants, who were unable to work due to ill health or were retired, taking part in OHCs could work as a replacement of role:

> *I feel that I can still use the knowledge and skills from my previous work along with my personal experience of asthma.* [N.3]
>
> *I use it far more when I am unwell with asthma.* [N.4]

Financial or Social Recognition Not Important

A question addressed whether the contribution superusers make could be recognized in any way (socially or financially), considering it might help other patients to manage their illness better. Of the 17 participants, 13 replied negatively, 3 were unsure, and only 1 replied positively.

The main reason behind the *no* answer was that reward should come from the awareness of helping others and the fact that social interaction is actually enjoyable.

The motivation to ensure that all users felt equally important to the whole community also played a role:

> *There is no guarantee that a superuser is any better informed than any other user. Superuser status might make others feel their contributions were not worth making. In my view, the forum should be one where every voice has equal status.* [N.9]

Other factors against social and financial recognition were the voluntary nature of contributions and the potential misuse of financial incentives:

> *I do it because I can, if I didn’t want to I wouldn't.* [N.16]

Some positively saw the social recognition as sensible/knowledgeable users:

> *Those who are particularly helpful could be recognised with titles, so users know who to contact for issues. Monetary incentives could also be considered, but most people who help online are doing it just to be helpful rather than for any other purposes.* [N.10]

Awareness of the Limits of Peer Self-Management Support

Decisions on Posts to Respond to

When asked about what determines their decision to reply to certain posts, participants showed a reassuring awareness of the type of self-management support they were able to offer (ie, emotional and behavioral but not medical tasks):

> *By drawing in on your own experiences. You soon recognise familiarities with symptoms.* [N.1]
>
> *I choose the ones which I have some relevant knowledge or experience to reply to.* [N.3]

Some mentioned that they posted replies when they felt they were able to provide a different/unrepresented point of view with respect to the ones already given, which in turn could help others make decisions:

> *...there may not be another voice in that comment section giving the view I feel, so I may choose to add it.* [N.16]

Participants’ aim was to empower patients and carers through their own experiences:

> *I will post a reply from my own experience of asthma gained over 50 years.* [N.11]

Types of Support

The type of support most frequently provided by our respondents was mainly behavioral and emotional. In addition, most participants also mentioned their role in signposting users to source of information and support:

> *Reassurance, information, sharing my experience of a particular health issue, information about where to go for further advice and information.* [N.3]

Medical Self-Management Needing Health Care Professionals’ Input

Medical self-management was unanimously agreed upon as something that required consultation with HCPs, and all superusers had prior experience of referring other community members to their HCPs:

> *I never give specific medical advice though.* [N.10]
>
> *Almost every time, the default answer is always to contact your own medical help for guidance.* [N.16]

Challenges

Problems and Difficulties

Of the 17 superusers, 9 described problems and difficulties associated with their role in the OHCs (2 were unsure about it, 4 replied no, and 2 did not reply).

The main difficulty described by superusers was the worry they felt regarding other community members who were not successfully managing their asthma and not seeking appropriate medical help:

> *Members who put their health at risk by not realising how dangerous a situation they are in.* [N.17]

Other problems described included dealing with misunderstandings, spam, or posts promoting miracle cures or dangerous ideas (eg, buying medicine over the internet). Of the 17 participants, 9 had experience of reporting such posts to moderators:

> *Spam e mails, folks responding who’ve not understood my posts, prolonged communication.* [N.2]
>
> *People offering “miracle” cures; people not being supportive; going off topic of the original post.* [N.14]

Some users found it difficult to deal with the negative tone of some conversations, when the underlying aim was to complain:

> *Some people don't want to take advice and will just complain constantly no matter what you suggest.* [N.15]

Only 1 user mentioned being trolled once in the past and this being a negative experience. Asthma UK HealthUnlocked community was described as a *good forum*.

Posts Causing Superusers Worries/Stress

Posts causing superusers worry were about religion-based advocacies and derogatory or emotionally challenging stories. Posts offering bad advice, indicating that users had little knowledge of asthma and its gravity, and revealing a sense of responsibility of superusers to reply to posts and moral pressure toward other OHC users also caused worry and stress.

Moreover, superusers worried about posts from users who were struggling or acutely unwell and subsequently stopped posting or from users who had been chronically struggling with their asthma without seeking professional help:

> *Those that encourage people with less knowledge to take actions that could put their child in danger/people who are disparaging someone else's intelligence or understanding of the illness.* [N.16]
>
> *[Posts suggesting that] children can be in respiratory distress.* [N.16]

Suggestions for Policies and Guidance

Of the 17 participants, 8 believed that more policies and guidance should be available for asthma OHCs (2 did not, 4 were unsure, and 3 did not answer the question). In particular, they felt that additional policies and guidance should be introduced on the rules for safe engagement with asthma OHCs and for clarifying when emergency medical advice is needed. Some participants did acknowledge that such policies were already in place, though not all users seemed to be aware of them. A suggestion was made for new users to be encouraged to passively engage and read posts before active engagement:

> *New members should be encouraged to read without contributing at first. I think all members do this instinctively anyway...joining a social group is rarely instant. Good sites include few risks, made safe by the site rules, moderators and experienced users.* [N.5]

One participant recommended having policies and guidance about buying medications over the internet.

A number of participants highlighted the importance of quick removal of clearly bad advice so as to develop patient confidence in participation:

> *Quick removal or "bad" posts (including spam). [Moderators to] Respond to occasional posts (especially if there's an argument going on in the feed).* [N.13]

Interface With Health Care Professionals

Health Care Professionals' Awareness of Engagement With Asthma Online Health Communities

Most participants' HCPs (10 out of 17) were not aware of superusers' engagement with asthma OHCs. Only 3 out 17 participants reported that their OHCs' involvement was known by their HCPs, whereas 4 out of 17 participants were unsure. Even when the HCPs were aware, this was because superusers mentioned their engagement with OHCs, though they did not discuss it any further:

> *The only person who knows is my husband.* [N.3]

Of the 17 participants, 15 stated that they did not believe that HCPs would discourage participation in asthma OHCs:

> *I think the quality of the BLF and Asthma UK sites is generally accepted. Medics know I'm am open minded and analytical about any information.* [N.5]

Only 1 participant reported being discouraged by their HCP from engaging with OHCs, and this was linked to concerns about patients becoming focused on illness rather than health and well-being:

> *They seemed to feel that by engaging with other people in online health forums it focused people on the illness rather than on getting on with life. They seemed to feel that it made people more anxious about their illness rather than provide reassurance, information and support. It seems to me that they were worried that it reinforced an 'illness' mentality. However, my experience is that generally it empowers people to take control of their own lives and make decisions in partnership with their medical carers rather than feel disempowered and uncertain.* [N.3]

Health Care Professionals Promote Engagement With Asthma Online Health Communities

The majority of participants (11 out 17) thought that HCPs should direct patients with LTCs to OHCs, provided they were appropriately moderated and trusted platforms. The remaining 6 out of 17 participants were unsure, though no one felt that HCPs should not promote engagement with OHCs:

> *Any recommended communities would need to be appropriately vetted/ endorsed by medical professional to ensure their accuracy in terms of medical advice and to keep people safe.* [N.10]

Respondents offered several specific suggestions about how to promote engagement with OHCs:

> *There could be posters up in the waiting rooms of relevant hospital departments and GP surgeries. Asthma nurses could inform patients. Ask people how they feel they have benefited from online communities.* [N.9]

Indeed, a range of advantages arising from the promotion of OHCs by HCPs included obtaining behavioral and emotional self-management support that HCPs may not be able to offer as easily:

> *For support for people when they get diagnosed, have a really difficult time with their asthma and recovering.* [N.11]

Suggestions to Reassure Health Care Professionals About Online Health Community Engagement

To reassure HCPs about the safety of OHCs, participants felt OHC providers should have clearer statements about contacting HCPs for medical self-management, place more emphasis on the fact that posts from peers come from not medically qualified people, and have a readily accessible guidance about keeping safe in social media. Comments from moderators should be regular and nonintrusive, with strict rules regarding posts:

> *Healthcare professionals may need to be reassured that any group they signpost is a medically sound one. However, they cannot dictate. It is about mutual respect for the role of the medical professional and the role of an on-line health community.* [N.3]

Participants felt improving HCPs' knowledge and awareness of why patients engage with OHCs and the benefits of peer support on LTCs would make them keener to promote OHCs. Evaluation of the impact of engagement in OHCs on patients was also suggested:

> *[HCPs' awareness that] online communities are primarily useful for feeling more "normal" with your condition - connecting with others in the same situation.* [N.10]

Health Care Professionals' Participation in Online Health Communities

When exploring whether HCPs should themselves take part in OHCs, 9 out of 17 participants replied positively, 5 were unsure, 2 were against it, and 1 did not answer the question.

The reasons behind perceiving HCPs' participation beneficial were the opportunity to get worries and questions addressed. However, as this respondent notes, their participation may be mutually beneficial through learning more about the patient experience of their illness:

> *Not only could a lot of peoples' worries and questions be easily answered authoritatively healthcare professionals could gain much knowledge from forums.* [N.8]

There was a mention of engagement in OHCs as an additional remunerated duty for HCPs:

> *I think they [HCPs] should be paid to set aside time to monitor forums.* [N.8]

Most participants felt that HCPs' participation in OHCs was important as long as their identity was stated:

> *They should include their medical specialisms in their profiles and understand that there are many viewpoints on some issues.* [N.5]

Difficulties making participants unsure about HCPs' participation were potential scrutiny of all posts, limitation of expression from different points of view, and the problem of not knowing the clinical details of users well enough before an appropriate answer could be given:

> *I feel the community is for those who don't otherwise have a voice and that it would seem too "preachy" to have a medic commenting on every post.* [N.9]
>
> *...it might get difficult (difficult/awkward for medics when dealing with people they know little about).* [N.13]

Issues with HCPs' code of conduct and difficulties with HCPs being patients themselves were also expressed:

> *Difficult. There is a place for it but I think it blurs the lines a little and their code of conduct with their registering body...I think as a healthcare professional who is also a patient they need to be aware of the blurred line between patient and healthcare worker.* [N.15]

Discussion

Principal Findings

This is the first study to provide evidence of superusers' motivations for engagement in a large nationwide OHC, the challenges they face when interacting with other users, and their interface with HCPs. As the use of social media in health care is increasing, taken together with our previous network study [25], these results provide unprecedented insight on superusers who are key to creating value, driving and sustaining user engagement and contributing to the success of an OHC.

Superusers are both patients with asthma and carers of a wide age range, tend to take part in more than one OHC, and spend considerable time in a role sometimes similar to that of moderators [37]. Reassuringly, they showed awareness of the complexity of coping with asthma and the limits of their advice, provided emotional and behavioral self-management support, and had at times to direct users to HCPs for medical queries. This is an important point as much of the work exploring HCPs' views of OHCs suggests that they are concerned that inappropriate advice is commonly shared and that community members may not be skilled/reflective enough to realize it.

The superuser role appears to be acquired by users as they deepen their asthma-related knowledge and become accustomed to web-based communication and the dynamics of group-based anonymous interaction [26], turning into *expert patients*, acquiring some of the characteristics of the *second generation of e-patients* [8].

Although the superuser role could be stressful at times, most HCPs were unaware of superusers' engagement with OHCs and therefore unable to provide support. This is also in contrast with the general agreement among superusers that patient engagement with trusted and thriving OHCs should be promoted within health care. For some, being a superuser could work as a *replacement* of role, as in the case of a retired HCP participant or a participant in working age who is identified as an HCP off work due to asthma.

Superusers who were themselves HCPs raised the issue of the need to develop a code of conduct within their registering bodies to engage with users in OHCs.

It has been suggested that HCPs' engagement with OHCs could be remunerated as part of HCP duties.

Superusers' perspectives on what would make OHCs safer and more effective are of interest not only to OHC platform providers but also to policymakers who are increasingly considering leveraging OHCs for health care delivery.

Strengths and Limitations

There are a number of strengths and limitations to our work, which merit comment. The data we collected from superusers in this study came from an existing and thriving asthma OHC [25,27]. In our previous study, we uncovered the emergence of superusers (or hubs) as the OHC network grew larger. Our findings suggest that users with a disproportionate number of contacts started to emerge only when many users had already joined the network (about 1000). This has important implications for the size of our sample of superusers. Although the absolute number of superusers in this study may appear to be small, it takes a very large-scale network for these superusers to emerge. Thus, our sample size must be gauged jointly with the (large) size of the underlying network to which the superusers belong. Although no saturation criteria were used to determine the study sample size, our qualitative analysis revealed that saturation of emerging themes was reached.

The currently limited literature about superusers in OHCs, the lack of a formal identification of *superuser status* in OHCs, and the a posteriori definition of superusers (ie, superusers as the top 1% active users over a 10-year period [25]) make it difficult to judge the response rate in this study. Our previous work [25] showed the presence of 20 to 30 superusers active on a weekly basis. Although for obvious reasons we could not use the same definition to identify superusers in this study, based on our previous results, the superuser response rate would support the validity of the data presented here.

The study benefited from a superuser piloting phase that face-validated the questions and improved their focus, resulting in additional questions. The study was not designed to test the self-determination theory, which was used as an interpretive lens.

The Asthma UK and Facebook communities are established OHCs (Asthma UK OHC has been operational since 2006) and are moderated and trusted; thus, the results may not extend to other OHCs. Although we cannot confirm superuser sharing of scientific information being always appropriate, in such circumstances, it is likely moderators and other superusers, as seen in this study, would intervene in providing rectification.

Moreover, the self-selective nature of recruitment may have introduced a subjective bias, as less altruistic superusers with different characteristics may not have responded to the invitation.

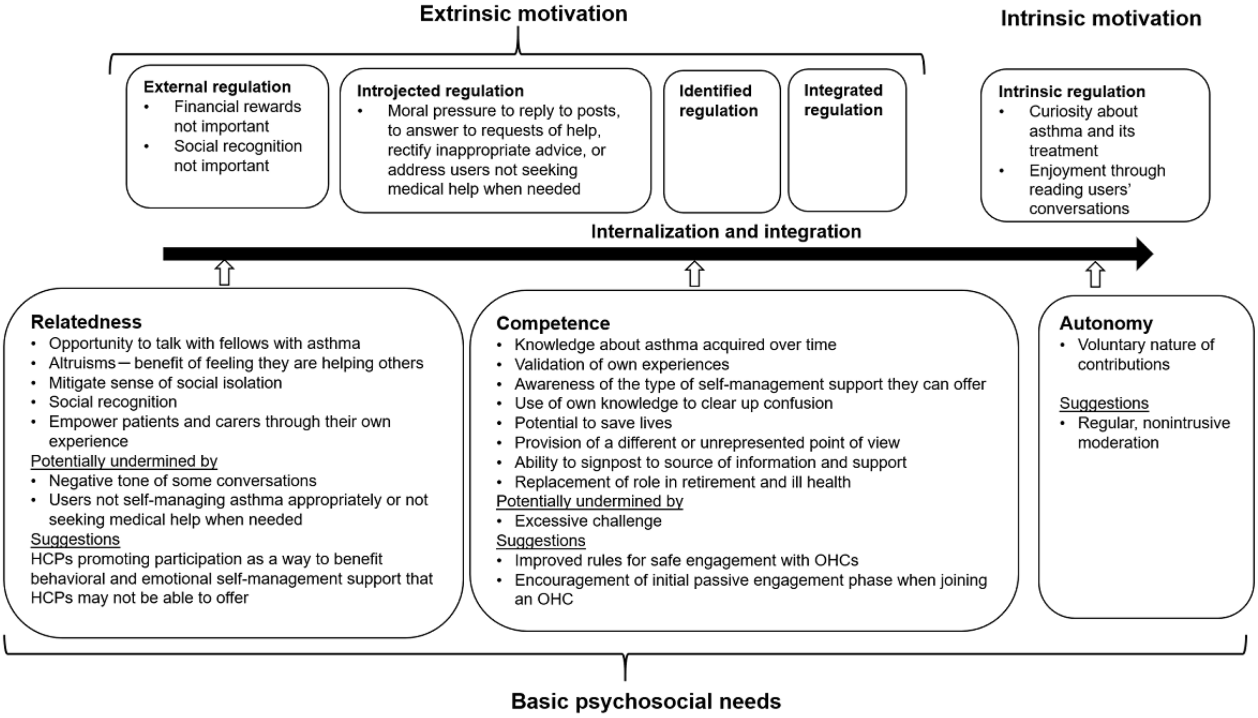
Comparison With Existing Literature

Only a handful of studies characterizing superusers in OHCs are present in the literature, and to our knowledge, this is the first direct account of superusers' motivation to engage in OHCs. Previous work focused on quantifying superusers [21] and their posting behavior [24] using a passive approach. Superusers have been described as mostly female [24], at times, assuming a role similar to moderators [37]. The desire for agency and mastery in asthma patients has been previously described [38]. Engagement with OHCs promoted superusers' sense of personal control/agency/self-efficacy over their illness, particularly when adopting the informal role of wise mentors or supporters to other users. Interestingly, a recent study indicated that patients gained empowerment through OHCs, which was positively related to patient commitment to the physician and to patient compliance with the proposed treatment [2]. Moreover, there is evidence that users who are high engagers (such as superusers who are themselves patients) exhibit the greatest improvement in patient activation measure (PAM; a measure that captures the extent to which people feel engaged and confident in taking care of their condition) in HealthUnlocked OHCs, even if the average change in PAM across all levels of engagement is not clinically meaningful [39].

Interpretation of Findings Through the Lens of Self-Determination Theory

Superusers display high intrinsic motivation to engage with OHCs (Figure 1) [31]. Intrinsically motivated behaviors are carried out for the sake of sheer interest or satisfaction derived from the task. Intrinsic motivation constitutes the most autonomous form of motivation and is highly evident in the participants of this study [31]. Through OHCs' engagement, they exhibited fulfillment of the 3 basic psychosocial needs: relatedness, competence, and autonomy. With respect to relatedness, superusers described a sense of belonging to the community and a feeling that they mattered to other users. Participants also expressed a sense of mastery (competence), believing in the effectiveness of their ongoing interactions with users within the OHCs. Their behavior is self-endorsed, reflecting autonomy. Superusers are autonomous and wholeheartedly behind their engagement with OHCs. With such strong intrinsic motivation, extrinsic motivation, that is, behaviors that are carried out to obtain outcomes unrelated to the activity itself, such as financial rewards, unsurprisingly, is not particularly relevant. Nevertheless, moral pressure to monitor OHCs, answer to requests of help, rectify any inappropriate advice, or address users not seeking medical help when appropriate were extrinsic motivation factors that at times felt difficult and stressful, needing to be internalized and integrated in their role of superusers.

Figure 1. Superusers' self-determination theory, freely adapted from Ryan and Deci's theory. Intrinsic motivation constitutes the most autonomous form of motivation and is highly evident in superusers. Such motivation emerges from pure personal interest, curiosity, or enjoyment through engagement with online health communities. The transition from external to intrinsic regulation is promoted by superusers' fulfillment of the 3 basic psychosocial needs: relatedness, competence, and autonomy. Within the basic psychosocial needs, factors potentially undermining fulfillment and suggestions for improvement are listed. HCP: health care professional; OHC: online health community.



Clinical and Research Implications

There is a need to improve clinicians,' researchers,' and policymakers' awareness of superusers. Clinicians could inquire about OHCs' engagement during consultations with patients with LTCs and offer support to any potential superusers. The first step is establishing a definitive trial to determine whether a primary care intervention specifically aimed at promoting engagement with trusted and thriving OHCs improves the health and well-being of patients with LTCs.

If integration of OHCs proves to be beneficial, given superusers' potential (ie, 10 superusers can sustain a community of 1000 people) [25,27], campaigns to promote patients with LTCs actively engaging with disease-specific and trusted OHCs may be a way to tackle the demand for behavioral and emotional self-management support in LTCs. Through participation in OHCs, patients who were unable to work due to ill health or retired naturally acquire over time the role of superusers and become a resource to the community, as shown in this study.

Further research should investigate the possible role of HCPs in OHCs based on their monitoring activity and contributions to web-based conversations. Indeed, the Big White Wall [15] and Health Service Executive Facebook [16] are already including HCPs in the delivery of health care.

Further studies of OHC superusers are needed, aimed at addressing their unmet needs and understanding their role as mentors, their learning potential, and how other users within the community learn from them. Using more explicitly self-determination theory approaches can inform the design of new theoretically informed strategies for planning, managing, and sustaining OHCs.

As with the UK NHS face-to-face peer supporters in mental health, the usefulness and development of potential training packages for superusers could be explored.

Superusers expressed the need to improve OHC moderation through quicker removal of harmful posts. This could be achieved by taking advantage of advances in artificial intelligence, which increasingly allow real-time monitoring of OHCs and identifying and *quarantining* posts until review by moderators.

HCPs' registering bodies may need to develop a code of conduct for HCPs' participation in OHCs, especially when they take on a superuser role.

These results should be considered in the current increasingly wider uptake of digital skills across populations, with 95% of UK adults being on the internet [40]. In this context, OHCs assume growing potential as vehicles of health and social interventions [41], with the presence of superusers playing a key role in guaranteeing OHC success or failure. The roll out of the NHS app through the National Health Service [42] and initiatives such as the Online Centres Network [43] are working to tackle digital and social exclusion by providing people with the skills and confidence they need to access digital technology. Indeed, 70% of homeless people use social media [44], and the estimated penetration of broadband connection ownership and the tendency to be influenced by web-based content are wider in ethnic minorities [45].

This study offers a novel and fresh perspective on motivation, difficulties, and interaction with HCPs of superusers, a group of patients likely to be key players in the digital health social media landscape.

Acknowledgments

AD was funded through the Barts Charity *real health* grant. The study is based on a previous work funded by a Queen Mary University of London Life Science Initiative grant (supported by the Wellcome Trust Institutional Strategic Support Fund). AS received support from Asthma UK for the Asthma UK Centre for Applied Research and from the Medical Research Council and other funders for the Health Data Research UK.

Conflicts of Interest

JP works as the Research Operations Officer at Asthma UK. AS acknowledged research grants from Asthma UK and Health Data Research UK. OF is a Patient Partner in the Patient Centricity Platform at Astra Zeneca. She is also a writer for the TEVA Life Effects web platform. She contributes to The Mighty, WEGO Health and a variety of online health platforms. She is a Research and Policy Volunteer with Asthma UK and part of the European Patient Pharma Forum.

References

1. Allen C, Vassilev I, Kennedy A, Rogers A. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
2. Audrain-Pontevia A, Menvielle L, Ertz M. Effects of three antecedents of patient compliance for users of peer-to-peer online health communities: cross-sectional study. J Med Internet Res 2019 Nov 11;21(11):e14006 [FREE Full text] [doi: 10.2196/14006] [Medline: 31710295]
3. National Voices. 2015. Peer Support: What Is It and Does It Work? A Summary of the Evidence URL: https://www.nesta.org.uk/sites/default/files/peer_support_-_what_is_it_and_does_it_work.pdf [accessed 2018-05-13]
4. Mo PK, Coulson NS. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459. [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
5. Pendry LF, Salvatore J. Individual and social benefits of online discussion forums. Comput Hum Behav 2015 Sep;50:211-220. [doi: 10.1016/j.chb.2015.03.067]
6. Bartlett YK, Coulson NS. An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication. Patient Educ Couns 2011 Apr;83(1):113-119. [doi: 10.1016/j.pec.2010.05.029] [Medline: 20599338]
7. Izuka NJ, Alexander MA, Balasooriya-Smeekens C, Mant J, De Simoni A. How do stroke survivors and their carers use practitioners' advice on secondary prevention medications? Qualitative study of an online forum. Fam Pract 2017 Sep 1;34(5):612-620. [doi: 10.1093/fampra/cmx023] [Medline: 28398553]
8. Duncan TD, Riggare S, Koch S, Sharp L, Hägglund M. From information seekers to innovators: qualitative analysis describing experiences of the second generation of e-patients. J Med Internet Res 2019 Aug 15;21(8):e13022 [FREE Full text] [doi: 10.2196/13022] [Medline: 31418421]
9. Hobbs WR, Burke M, Christakis NA, Fowler JH. Online social integration is associated with reduced mortality risk. Proc Natl Acad Sci U S A 2016 Nov 15;113(46):12980-12984 [FREE Full text] [doi: 10.1073/pnas.1605554113] [Medline: 27799553]
10. Panagioti M, Richardson G, Small N, Murray E, Rogers A, Kennedy A, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. BMC Health Serv Res 2014 Aug 27;14:356 [FREE Full text] [doi: 10.1186/1472-6963-14-356] [Medline: 25164529]
11. Taylor S, Pinnock H, Epiphanou E, Pearce G, Parke H, Schwappach A. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS - practical systematic review of self-management support for long-term conditions. Health Serv Deliv Res 2014;2(53):-. [doi: 10.3310/hsdr02530] [Medline: 25642548]
12. De Simoni A, Griffiths CJ, Taylor SJ. Improving access to primary care: can online communities contribute? Br J Gen Pract 2016 Nov;66(652):559 [FREE Full text] [doi: 10.3399/bjgp16X687721] [Medline: 27789494]
13. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
14. Taylor J, Pagliari C. The social dynamics of lung cancer talk on Twitter, Facebook and Macmillan.org.uk. NPJ Digit Med 2019;2:51 [FREE Full text] [doi: 10.1038/s41746-019-0124-y] [Medline: 31304397]
15. Harding C, Chung H. Behavioral health support and online peer communities: international experiences. Mhealth 2016;2:43 [FREE Full text] [doi: 10.21037/mhealth.2016.10.04] [Medline: 28293613]

16. Facebook. You Can Quit URL: https://www.facebook.com/pg/HSEquit/about/?ref=page_internal [accessed 2020-05-28]
17. Facebook. Stoptober URL: https://www.facebook.com/pg/stoptober/about/?ref=page_internal [accessed 2020-05-28]
18. NHS Digital. 2019. New Social Media Techniques Used to Boost Cancer Screening Rates URL: https://digital.nhs.uk/news-and-events/latest-news/new-social-media-techniques-used-to-boost-cancer-screening-rates [accessed 2020-05-28]
19. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new medical research council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
20. Bender JL, Jimenez-Marroquin MC, Ferris LE, Katz J, Jadad AR. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. Support Care Cancer 2013 May;21(5):1253-1263. [doi: 10.1007/s00520-012-1655-9] [Medline: 23179491]
21. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014 Feb 4;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
22. Cunningham JA, van Mierlo T, Fournier R. An online support group for problem drinkers: AlcoholHelpCenter.net. Patient Educ Couns 2008 Feb;70(2):193-198. [doi: 10.1016/j.pec.2007.10.003] [Medline: 18022340]
23. Corbin J, Strauss A. Managing chronic illness at home: three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/bf00989485]
24. van Mierlo T, Voci S, Lee S, Fournier R, Selby P. Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian cancer society's smokers' helpline online and StopSmokingCenter.net. J Med Internet Res 2012 Jun 26;14(3):e66 [FREE Full text] [doi: 10.2196/jmir.1854] [Medline: 22732103]
25. Joglekar S, Sastry N, Coulson NS, Taylor SJ, Patel A, Duschinsky R, et al. How online communities of people with long-term conditions function and evolve: network analysis of the structure and dynamics of the asthma UK and British lung foundation online communities. J Med Internet Res 2018 Jul 11;20(7):e238 [FREE Full text] [doi: 10.2196/jmir.9952] [Medline: 29997105]
26. McDonald D, Woodward-Kron R. Member roles and identities in online support groups: perspectives from corpus and systemic functional linguistics. Discourse Commun 2016 Jan 25;10(2):157-175 [FREE Full text] [doi: 10.1177/1750481315615985]
27. De Simoni A, Taylor S, Griffiths C, Panzarasa P, Sheikh A. NEJM Catalyst. 2018. Online 'Superusers' as Allies of the Health Care Workforce URL: https://catalyst.nejm.org/doi/abs/10.1056/CAT.18.0052 [accessed 2020-05-28]
28. Selby P, van Mierlo T, Voci SC, Parent D, Cunningham JA. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010 Aug 18;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]
29. Panzarasa P, Grffiths CJ, Sastry N, De Simoni A. Social medical capital: how patients and caregivers can benefit from online social interactions. J Med Interent Res 2020:- (forthcoming). [doi: 10.2196/16337]
30. van Mierlo T, Hyatt D, Ching AT. Mapping power law distributions in digital health social networks: methods, interpretations, and practical implications. J Med Internet Res 2015 Jun 25;17(6):e160 [FREE Full text] [doi: 10.2196/jmir.4297] [Medline: 26111790]
31. Ryan R, Deci E. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
32. De Simoni A, Horne R, Fleming L, Bush A, Griffiths C. What do adolescents with asthma really think about adherence to inhalers? Insights from a qualitative analysis of a UK online forum. BMJ Open 2017 Jun 13;7(6):e015245 [FREE Full text] [doi: 10.1136/bmjopen-2016-015245] [Medline: 28615272]
33. De Simoni A, Shanks A, Balasooriya-Smeekens C, Mant J. Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative study of a sample of participants. BMJ Open 2016 Apr 6;6(4):e010501 [FREE Full text] [doi: 10.1136/bmjopen-2015-010501] [Medline: 27053271]
34. SurveyMonkey. 2018. Research Study: Why Do Some Members Engage Much More in an Online Health Community? URL: https://www.surveymonkey.co.uk/r/CP7JQ7L [accessed 2020-05-28]
35. Asthma UK Centre for Applied Research. Public Involvement URL: https://www.aukcar.ac.uk/public-involvement [accessed 2020-05-28]
36. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
37. Coulson N, Shaw R. Nurturing health-related online support groups: exploring the experiences of patient moderators. Comput Hum Behav 2013 Jul;29(4):1695-1701 [FREE Full text] [doi: 10.1016/j.chb.2013.02.003]
38. Pinnock H, Slack R, Pagliari C, Price D, Sheikh A. Understanding the potential role of mobile phone-based monitoring on asthma self-management: qualitative study. Clin Exp Allergy 2007 May;37(5):794-802. [doi: 10.1111/j.1365-2222.2007.02708.x] [Medline: 17456228]

39. Costello RE, Anand A, Jameson Evans M, Dixon WG. Associations between engagement with an online health community and changes in patient activation and health care utilization: longitudinal web-based survey. J Med Internet Res 2019 Aug 29;21(8):e13477 [FREE Full text] [doi: 10.2196/13477] [Medline: 31469082]
40. Prescott C. Office for National Statistics. 2019. Internet Users, UK: 2019 URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2019 [accessed 2019-11-01]
41. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun 25;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
42. NHS Digital. 2020. NHS App URL: https://digital.nhs.uk/services/nhs-app [accessed 2020-02-01]
43. Online Centres Network. 2020. URL: https://www.onlinecentresnetwork.org/ [accessed 2020-02-01]
44. Guadagno R, Muscanell N, Pollio D. The homeless use Facebook?! Similarities of social network use between college students and homeless young adults. Comput Hum Behav 2013 Jan;29(1):86-89 [FREE Full text] [doi: 10.1016/j.chb.2012.07.019]
45. OfCom. 2013. Ethnic Minority Groups and Communication Services URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/62870/ethnic-minority-groups.pdf [accessed 2020-05-28]

Abbreviations

BLF: British Lung Foundation
HCP: health care professional
NHS: National Health System
LTCs: long-term conditions
OHC: online health community
PAM: patient activation measure
PPI: Patient and Public Involvement

Edited by G Eysenbach; submitted 10.02.20; peer-reviewed by T Scott Duncan, D Hyatt; comments to author 27.02.20; revised version received 23.04.20; accepted 13.05.20; published 23.06.20.

Please cite as:
De Simoni A, Shah AT, Fulton O, Parkinson J, Sheikh A, Panzarasa P, Pagliari C, Coulson NS, Griffiths CJ
Superusers' Engagement in Asthma Online Communities: Asynchronous Web-Based Interview Study
J Med Internet Res 2020;22(6):e18185
URL: http://www.jmir.org/2020/6/e18185/
doi:10.2196/18185
PMID:32573463

Original Paper

Examining Social Capital, Social Support, and Language Use in an Online Depression Forum: Social Network and Content Analysis

Wenjing Pan[1], PhD; Bo Feng[2], PhD; Cuihua Shen[2], PhD

[1]School of Journalism and Communication, Renmin University of China, Beijing, China

[2]Department of Communication, University of California, Davis, Davis, CA, United States

Corresponding Author:
Wenjing Pan, PhD
School of Journalism and Communication
Renmin University of China
507, School of Journalism and Communication
Renmin University of China, 59 Zhongguancun St, Haidian District
Beijing, 100872
China
Phone: 86 010 82500855
Email: wenjingpan@ruc.edu.cn

Abstract

Background: The use of peer-to-peer online support groups and communities has grown into a social phenomenon. Many people use online support groups and communities to seek and provide social support. It is essential to examine how users' participation behaviors may contribute to different outcomes.

Objective: This study aimed to (1) use the structural positions of online depression forum users in their reply network to predict received support and (2) examine their language use reflecting their health conditions.

Methods: A total of 2061 users and their 62,274 replies posted on a depression forum from July 2004 to July 2014 were extracted using a web crawler written in Python. The content of the forum users' posts and replies and their reply patterns were examined. A social network analysis method was used to build the reply networks of users. The computerized text analysis method was used to measure features of the forum users' language styles.

Results: Forum users' bridging social capital (operationalized as network betweenness) was positively associated with the level of communication accommodation in their received replies (*P*=.04). Forum users' bonding social capital (operationalized as network constraint) was negatively associated with the level of communication accommodation in their received replies (*P*<.001). The forum users' change in their use of self-referent words and words expressing negative emotions were examined as linguistic proxies for their health conditions and mental states. The results revealed a general negative association between the number of received replies and the degree of decrease in the use of words expressing negative emotion (*P*=.007).

Conclusions: The structural positions of online depression forum users in the reply network are associated with different participation outcomes in the users. Thus, receiving replies can be beneficial to online depression forum users.

(J Med Internet Res 2020;22(6):e17365) doi:10.2196/17365

KEYWORDS

social capital; social support; social network analysis; computerized text analysis; communication accommodation; language style matching; online support forums; depression; mental health

Introduction

Background

Social capital, broadly defined as the potential benefits one receives from relationships with others, is a well-researched concept in many social science disciplines. Social capital has been associated with various positive outcomes at both the individual level (such as an increased sense of trust and belonging and improved mental and physical health) [1] and societal level (such as improved public health, promoted social integration, and decreased crime rate) [2]. With the popularity of internet technologies and social networking sites, a growing number of people are turning to web-based venues to seek social support. Many studies have documented that social media and online support groups generate social capital [3].

Social support is a concept closely related to the concept of social capital. Social support encompasses the comfort, assistance, and reassurance that people experience as a function of social relationships [4]. Social support can be viewed as a product resulting from the interpersonal relationships that constitute a support network. A unique feature of seeking and providing social support online is that individuals have more opportunities to communicate with others who share similar experiences and greater access to weak-tie networks [5]. The social capital and weak-tie network theories have also been adopted to explain why and how individuals may benefit differently from participating in online support groups [6,7].

Objectives

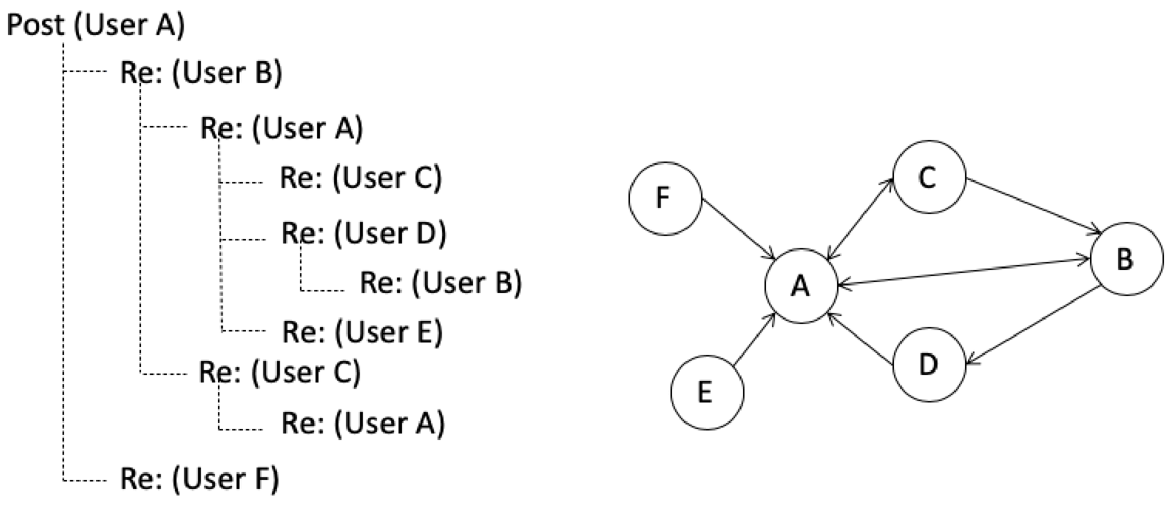
Previous studies investigating the network features of support seeking and provision in online support groups rarely measured the actual tie strength or examined the structural features of the support network. Furthermore, studies focusing on the content and qualities of support-seeking and support-provision messages are rare. Without examining the content of the exchanged messages, we cannot tell how support-seeking and support-provision messages are qualitatively different from each other. Utilizing big data–powered techniques such as web scraping and computerized text analysis, this study aimed to analyze the reply patterns and language styles of online depression forum users. The social capital and communication accommodation theories were adopted to make and test predictions about the association between social capital and social support of online support forum users. Drawing on the buffering effect model and self-awareness theory of depression, this study also examines the health benefits resulting from receiving social support by monitoring the changes in depression forum users' language use over time.

Theories and Hypotheses

Social Capital in Online Support Forums

Active participation in an online support forum generally takes one or both of the following 2 forms: contributing an original post and responding to posts by other users. Original posts typically reflect support-seeking efforts, whereas replying to the posts of other users can be seen as support-provision efforts [7]. Replies to an original post indicate attention, engagement, and responsiveness to the posts [8,9]. Therefore, an online support forum can be conceptualized as a network of users connected by their reply-based relationships (Figure 1). An online support forum provides social capital in that it is an embedded community activated for purposeful action [10]. Social support can be viewed as an outcome of social capital because the latter provides individuals with connections or relations to receive social support from network members when needed [11].

Figure 1. Illustration of a hypothetical user discussion network.



Bridging and Bonding Social Capital

Social capital can be categorized into two types: bridging and bonding social capital [12]. Bridging social capital is linked to weak ties, which are loose connections between individuals and are better for linkage to external assets and information diffusion [13]. Research has shown that bridging social capital may translate to job-related information [13], higher salaries and promotion opportunities [14], and better task performance in virtual worlds [15]. Bonding social capital is inward looking and tends to reinforce exclusive identities and homogeneous groups [12]. Bonding social capital has been shown to be associated with greater emotional support, better general health and well-being, and a greater level of social trust [16].

Communication Accommodation Reflecting Social Support

As human beings, we consciously and unconsciously match each other's communication acts (both verbally and nonverbally) [17]. From the message production end, communication accommodation is a socially motivated process where people engage in more or less accommodation based on their social assessment and social status of themselves and others [18]. From the message reception end, communication accommodation tends to be positively evaluated and can result in positive future intentions toward interactions with another person [18].

Individuals go to online support forums to seek and provide informational, emotional, and esteem support to each other. As these online forums are text based, various forms of support are sought and provided through text form. Communication accommodation reflected in the form of linguistic accommodation conveys a sense of caring and engagement. Linguistic accommodation can be captured through language style matching (LSM). LSM measures linguistic accommodation by taking into account the production of identical words and grammatical structures between the interactants and encompasses conversation level and synchrony on a turn-by-turn level [17]. Past research also shows that LSM among health bloggers and their readers contributes to the bloggers' perceptions of the availability of support from the readers [19]. In this study, the level of linguistic accommodation in the received replies was employed as a proxy for social support.

Social Capital and Communication Accommodation

The concept of social capital can be understood structurally in that social capital refers to the "extent or the intensity of associational links or activity" [16]. The structural perspective represents an objective view of the concept of social capital, and the tie strength or the structural positions can be assessed accurately using the social network analysis method. The nature of social capital entails that, social capital, as a form of capital, can be viewed as an investment in social relations with expected returns. Bridging and bonding social capital represents two unique ways individuals choose to invest in their social relations. The different positions forum users have in the reply network represent their structural aspects of social capital. Users who possess more bridging or bonding social capital are those who (1) occupy important positions in the users' reply social network and (2) share distinctive interaction patterns with other users. In online support forums, by investing their limited time and resources to manage their social relations, forum users expect returns in the form of social support. On the basis of the previous discussion, communication accommodation reflected in forum users' replies can serve as an indicator of social support. In online support forums, the bridging and bonding capital of users should be positively associated with received social support. The accumulated interaction patterns among users should be associated with social support reflected in other users' communication accommodation toward them. Therefore, based on discussion regarding the association between social capital, social support, and communication accommodation, the following hypotheses were proposed:

Hypothesis 1: Forum users' bridging social capital will be positively associated with the communication accommodation of others in their replies.

Hypothesis 2: Forum users' bonding social capital will be positively associated with the communication accommodation of others in their replies.

Owing to the weak nature of ties where bridging capital is accumulated, substantial emotional support or tangible support is less frequently observed compared with informational support [20]. As bonding social capital facilitates the exchange of emotional support, whereas bridging social capital facilitates the exchange of informational support, users will be more engaged in communicating with others whom they share strong ties with. The increased engagement between two communication partners also increases their degree of communication accommodation [21]. The association between users' bridging social capital and the degree of communication accommodation in received replies should be weaker compared with the association between users' bonding social capital and the degree of communication accommodation in received replies. Therefore, the following hypothesis was proposed:

Hypothesis 3: The association between forum users' bridging social capital and the level of communication accommodation in their received replies will be weaker compared with the association between forum users' bonding social capital and the level of communication accommodation in their received replies.

Health Outcomes of Receiving Replies

One of the most important benefits related to participation in online support groups is the improvement of the individuals' psychological and physical health [22]. The buffering effect model maintains that social support helps individuals buffer the negative impacts caused by stressful events [23]. Drawing on the buffering effect model, studies have found that receiving social support in an online support community was negatively associated with perceived stress [24] and positively associated with online network size and perceived social support [25]. Social support exchanged in online support groups also facilitates the management of health-related uncertainty [26] and promotes positive health outcomes [27].

However, previous studies examining health outcomes mostly adopted self-reported measures [25,28]. Self-reported data are usually accompanied by the inaccuracy of memory or recall. Examining the changes in the language use of online support forum users can avoid (1) sampling error by web scraping and analyzing all messages exchanged in online depression forums and (2) other limitations usually associated with survey methodologies, including interviewing effect and inaccuracy of memory or recall. Several linguistic features can serve as indicators of people's mental states [29]. Depressed individuals are more self-focused, express more negative emotions, and sometimes use more death-related words [30,31]. Research has shown that depressed patients tend to use more first-person singular pronouns and words expressing negative emotions than people who have never been depressed [32].

On the basis of previous theories regarding the buffering effect model and health benefit, it can be predicted that there should be a positive association between forum users' number of received replies and their health improvement as reflected in their language use. Therefore, we predict the following two hypotheses:

Hypothesis 4: Forum users’ number of received replies will be negatively associated with the change in their use of first-person singular pronouns.

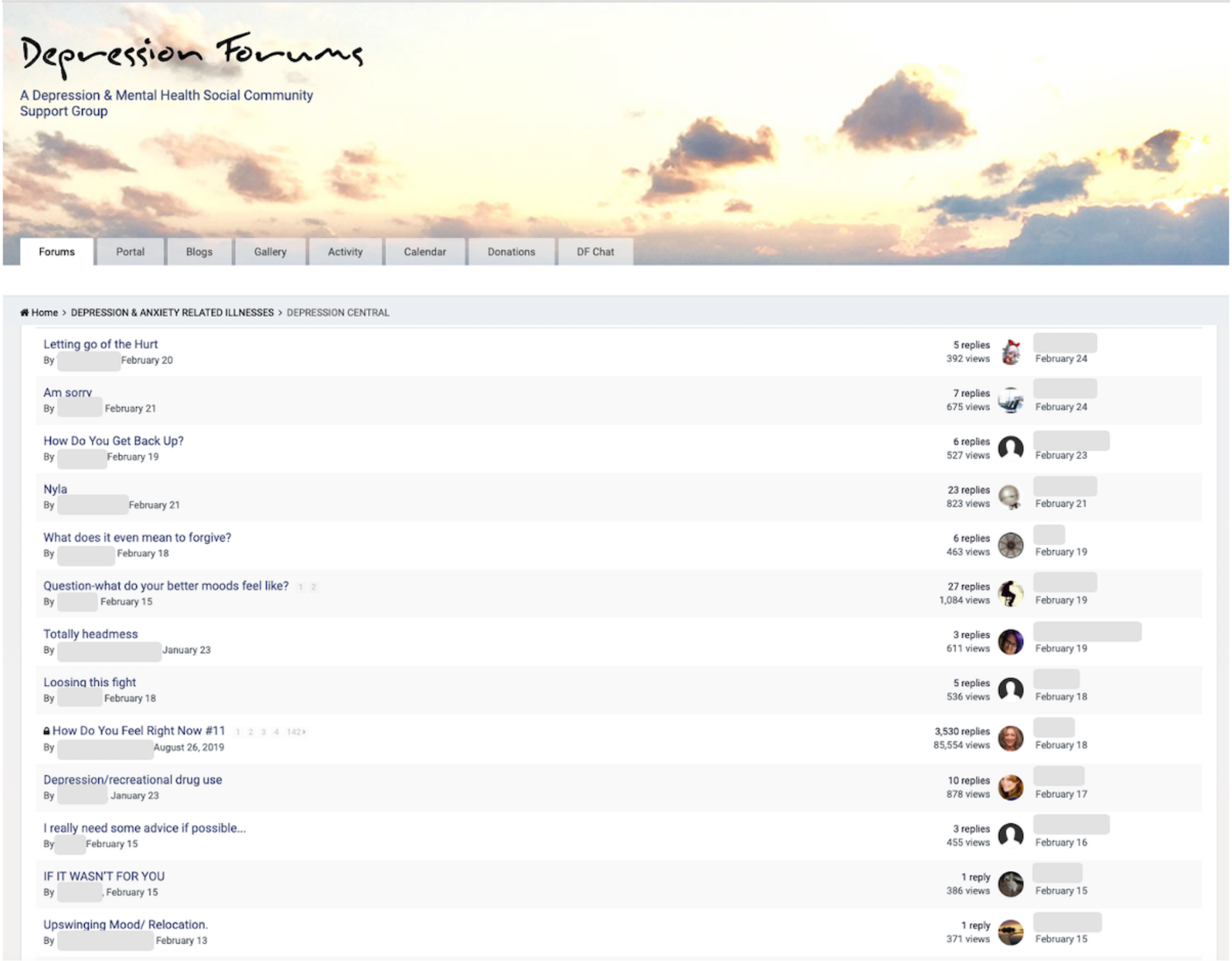
Hypothesis 5: Forum users’ number of received replies will be negatively associated with the change in their use of words expressing negative emotion.

Methods

Data Collection

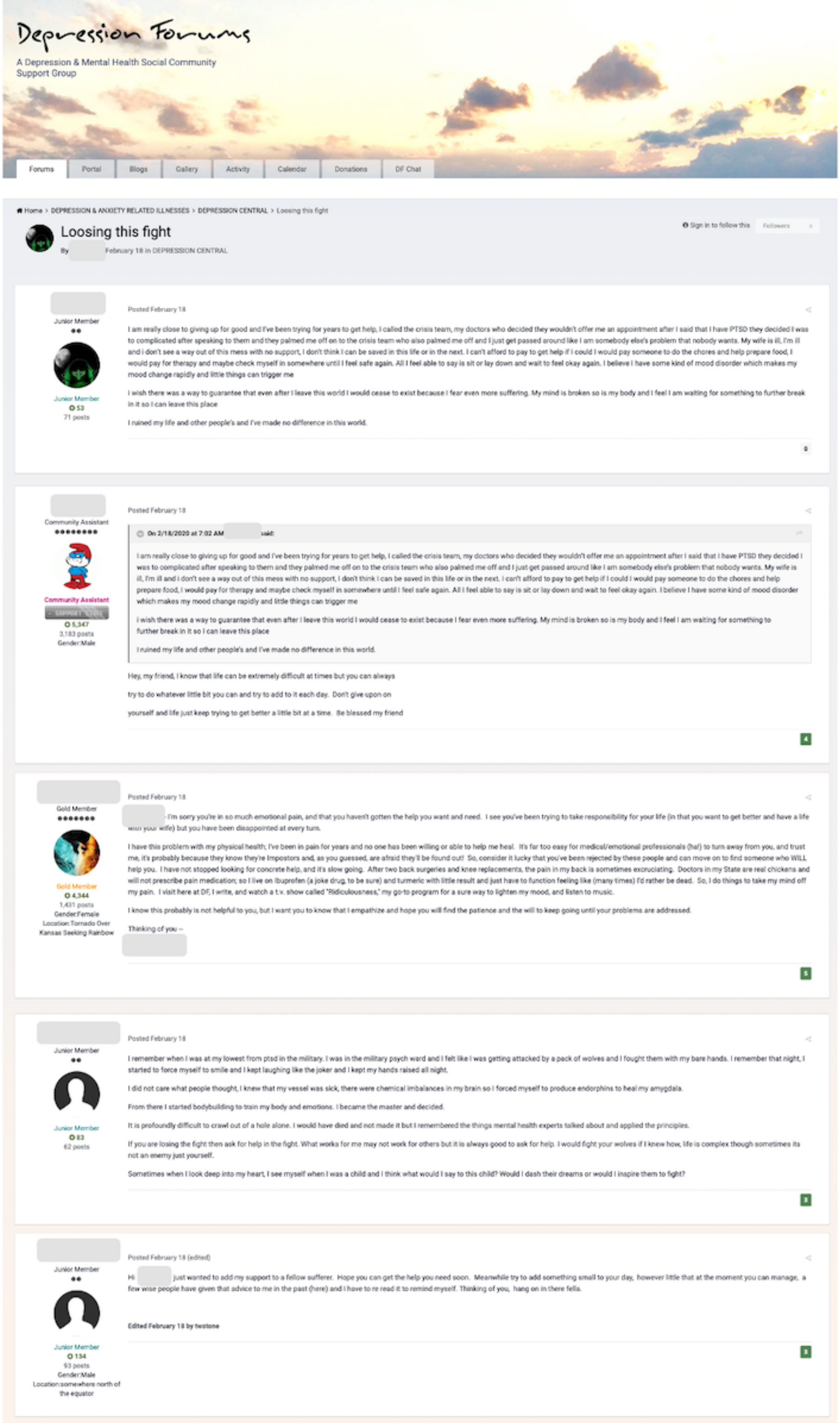
The depression forum examined in this study [33] is one of the largest peer-to-peer mental health community and support group in North America (see Figures 2 and 3 for screenshots). The forum is administered and moderated by volunteers rather than health professionals. Anyone who registers as a user of the forum can make a post or reply to others to participate in forum discussion.

Figure 2. Screenshot of the depression forum’s threads. The virtual identities of the forum users are masked.



Letting go of the Hurt By [redacted] February 20	5 replies 392 views	[redacted]	February 24
Am sorry By [redacted] February 21	7 replies 675 views	[redacted]	February 24
How Do You Get Back Up? By [redacted] February 19	6 replies 527 views	[redacted]	February 23
Nyla By [redacted] February 21	23 replies 823 views	[redacted]	February 21
What does it even mean to forgive? By [redacted] February 18	6 replies 463 views	[redacted]	February 19
Question-what do your better moods feel like? 1 2 By [redacted] February 15	27 replies 1,084 views	[redacted]	February 19
Totally headmess By [redacted] January 23	3 replies 611 views	[redacted]	February 19
Loosing this fight By [redacted] February 18	5 replies 536 views	[redacted]	February 18
How Do You Feel Right Now #11 1 2 3 4 142▸ By [redacted] August 26, 2019	3,530 replies 85,554 views	[redacted]	February 18
Depression/recreational drug use By [redacted] January 23	10 replies 878 views	[redacted]	February 17
I really need some advice if possible... By [redacted] February 15	3 replies 455 views	[redacted]	February 16
IF IT WASN'T FOR YOU By [redacted], February 15	1 reply 386 views	[redacted]	February 15
Upswinging Mood/ Relocation. By [redacted] February 13	1 reply 371 views	[redacted]	February 15

Figure 3. Screenshot of the depression forum's posts and replies. The virtual identities of the forum users are masked.



Depression Forums

A Depression & Mental Health Social Community Support Group

Home > DEPRESSION & ANXIETY RELATED ILLNESSES > DEPRESSION CENTRAL > Loosing this fight

Loosing this fight

By [redacted] February 18 in DEPRESSION CENTRAL

[redacted]
Junior Member
Junior Member
53
71 posts

Posted February 18

I am really close to giving up for good and I've been trying for years to get help, I called the crisis team, my doctors who decided they wouldn't offer me an appointment after I said that I have PTSD they decided I was to complicated after speaking to them and they palmed me off on to the crisis team who also palmed me off and I just get passed around like I am somebody else's problem that nobody wants. My wife is ill, I'm ill and i don't see a way out of this mess with no support, I don't think I can be saved in this life or in the next. I can't afford to pay to get help if I could I would pay someone to do the chores and help prepare food, I would pay for therapy and maybe check myself in somewhere until I feel safe again. All I feel able to say is sit or lay down and wait to feel okay again. I believe I have some kind of mood disorder which makes my mood change rapidly and little things can trigger me

I wish there was a way to guarantee that even after I leave this world I would cease to exist because I fear even more suffering. My mind is broken so is my body and I feel I am waiting for something to further break in it so I can leave this place

I ruined my life and other people's and I've made no difference in this world.

0

[redacted]
Community Assistant
Community Assistant
SUPPORT GUIDE
5,347
3,183 posts
Gender:Male

Posted February 18

> On 2/18/2020 at 7:02 AM [redacted] said:
>
> I am really close to giving up for good and I've been trying for years to get help, I called the crisis team, my doctors who decided they wouldn't offer me an appointment after I said that I have PTSD they decided I was to complicated after speaking to them and they palmed me off on to the crisis team who also palmed me off and I just get passed around like I am somebody else's problem that nobody wants. My wife is ill, I'm ill and i don't see a way out of this mess with no support, I don't think I can be saved in this life or in the next. I can't afford to pay to get help if I could I would pay someone to do the chores and help prepare food, I would pay for therapy and maybe check myself in somewhere until I feel safe again. All I feel able to say is sit or lay down and wait to feel okay again. I believe I have some kind of mood disorder which makes my mood change rapidly and little things can trigger me
>
> I wish there was a way to guarantee that even after I leave this world I would cease to exist because I fear even more suffering. My mind is broken so is my body and I feel I am waiting for something to further break in it so I can leave this place
>
> I ruined my life and other people's and I've made no difference in this world.

Hey, my friend, I know that life can be extremely difficult at times but you can always

try to do whatever little bit you can and try to add to it each day. Don't give upon on

yourself and life just keep trying to get better a little bit at a time. Be blessed my friend

4

[redacted]
Gold Member
Gold Member
4,344
1,431 posts
Gender:Female
Location:Tornado Over Kansas Seeking Rainbow

Posted February 18

[redacted] - I'm sorry you're in so much emotional pain, and that you haven't gotten the help you want and need. I see you've been trying to take responsibility for your life (in that you want to get better and have a life with your wife) but you have been disappointed at every turn.

I have this problem with my physical health; I've been in pain for years and no one has been willing or able to help me heal. It's far too easy for medical/emotional professionals (ha!) to turn away from you, and trust me, it's probably because they know they're Impostors and, as you guessed, are afraid they'll be found out! So, consider it lucky that you've been rejected by these people and can move on to find someone who WILL help you. I have not stopped looking for concrete help, and it's slow going. After two back surgeries and knee replacements, the pain in my back is sometimes excruciating. Doctors in my State are real chickens and will not prescribe pain medication; so I live on Ibuprofen (a joke drug, to be sure) and turmeric with little result and just have to function feeling like (many times) I'd rather be dead. So, I do things to take my mind off my pain. I visit here at DF, I write, and watch a t.v. show called "Ridiculousness," my go-to program for a sure way to lighten my mood, and listen to music.

I know this probably is not helpful to you, but I want you to know that I empathize and hope you will find the patience and the will to keep going until your problems are addressed.

Thinking of you --

[redacted]

5

[redacted]
Junior Member
Junior Member
83
62 posts

Posted February 18

I remember when I was at my lowest from ptsd in the military. I was in the military psych ward and I felt like I was getting attacked by a pack of wolves and I fought them with my bare hands. I remember that night, I started to force myself to smile and I kept laughing like the joker and I kept my hands raised all night.

I did not care what people thought, I knew that my vessel was sick, there were chemical imbalances in my brain so I forced myself to produce endorphins to heal my amygdala.

From there I started bodybuilding to train my body and emotions. I became the master and decided.

It is profoundly difficult to crawl out of a hole alone. I would have died and not made it but I remembered the things mental health experts talked about and applied the principles.

If you are losing the fight then ask for help in the fight. What works for me may not work for others but it is always good to ask for help. I would fight your wolves if I knew how, life is complex though sometimes its not an enemy just yourself.

Sometimes when I look deep into my heart, I see myself when I was a child and I think what would I say to this child? Would I dash their dreams or would I inspire them to fight?

3

[redacted]
Junior Member
Junior Member
134
93 posts
Gender:Male
Location:somewhere north of the equator

Posted February 18 (edited)

Hi [redacted] just wanted to add my support to a fellow sufferer. Hope you can get the help you need soon. Meanwhile try to add something small to your day, however little that at the moment you can manage, a few wise people have given that advice to me in the past (here) and I have to re read it to remind myself. Thinking of you, hang on in there fella.

Edited February 18 by twotone

3

Web-scraped data from the depression forum were used for this study [6]. The original forum was written in hypertext markup language in a structured format. The web crawler was written in the computer programing language, Python. Forum data including the title of the subforums, the title of the thread, post ID, reply ID, time stamps, full texts of the posts, and replies were extracted. The web-scraped data were then converted into JavaScript Object Notation (JSON) files. To perform a social network analysis, the adjacency matrix in comma separated value (csv) was constructed from the JSON file using the Python code. The Python code for the web crawler and network data extraction can be found in Multimedia Appendices 1 and 2.

The complete data set included 34,554 users and their 63,514 posts and 592,649 replies from July 2004 to July 2014. Users who never made any posts or replies (n=17,572) were removed from the analysis. The mean number of sent-out replies was 10 (SD 74.23), and the mean for the number of received replies was also 10 (SD 36.05). Furthermore, more than 94.04%

(32,493/34,553) of users had <10 sent-out and received replies combined over 10 years. Therefore, following a procedure similar to that of previous studies to capture users' active participation, only users who contributed ≥10 sent-out and received replies in total were included in the final data set [34,35]. The final data set included 2061 users and their 62,274 replies".

Data Analysis

On the basis of the final data set, an adjacency matrix was created to represent the users' reply patterns, with rows and columns representing each user and the number in each cell representing the number of replies between any given 2 users. Packages *sna* [36] and *network* [37] in the R programing language were used to calculate network-related variables such as the number of received replies, number of sent-out replies, betweenness, and constraint.

To calculate the forum users' language use styles, a computerized text analysis program, linguistic inquiry and word count (LIWC; Pennebaker, Booth, and Francis), was used [38]. LIWC is a dictionary-based software that compares and matches textual messages with predefined word categories. LIWC has been widely adopted and validated to examine the content and style features of various texts [30]. IBM's SPSS version 26 was used to perform linear regression analyses.

Measures

Bridging Social Capital

Following similar studies of online social capital [39], bridging social capital was structurally measured by each node's network betweenness. Betweenness measures the extent to which the actor falls on the geodesic (shortest) paths between other pairs of actors in the network [40]. Definitions of the key variables, operationalization, and measurements are included in Table 1.

Table 1. Definitions, operationalizations, and measurements of the key theoretical constructs.

Concepts	Definitions	Operationalizations	Measure
Bridging social capital	Bridging social capital is linked to weak ties, which are loose connections between individuals and are better for linkage to external assets and information diffusion	Network structural representation of structural hole or brokerage	Betweenness: Suppose that $g_i^{(st)}$ is the number of geodesic paths from actor *s* to actor *t* that pass through *I*, and that n_{st} is the total number of geodesic paths from *s* to *t*. The betweenness of the actor *i* can be given in the following formula [41]:
Bonding social capital	Bonding social capital is inward looking and tend to reinforce exclusive identities and homogeneous groups	Network structural representation of closely connected components	Constraint [42]: where p_{ij} is the proportion of ego *i*'s resource invested in connection with *j* and p_{qj} is the strength of *q*'s tie to *j*.
Social support	Social support encompasses the comfort, assistance, and reassurance that people experience as a function of social relationships	Communication accommodation reflected in forum users' posts and replies	Language style matching (LSM): For example, for personal pronouns (pp) between a given post of user A and the first received replies, the calculation would be:
Health conditions	Language use can reflect individuals' psychological states and health conditions	(1) Use of self-referent words (2) Use of words expressing negative emotion	(1) Change in the use of self-referent words=the number of self-referent words in last 10 replies−the number of self-referent words in the first 10 replies (2) Change in the use of words expressing negative emotion=the number of words expressing negative emotion in last 10 replies−the number of words expressing negative emotion in the first 10 replies

Bonding Social Capital

Bonding social capital was measured structurally as a network constraint. Constraint is "a concentration measure that varies from zero to 100 with the extent to which all of a person's network time and energy is concentrated in one contact" [42]. A large number of constraints suggest that the ego invests most resources within a small and connected group of nodes.

Language Style Matching

LSM scores were calculated for each user to measure the extent to which others have used similar language styles [19,43,44]. LSM was operationalized as the degree to which a user's post and received replies from others used a similar proportion of 9 classes of function words. Each forum user's use of 9 types of function words in each of their posts was calculated. The 9 types include auxiliary verbs (eg, to be, to have), articles (eg, an, the), common adverbs (eg, hardly, often), personal pronouns (eg, I, they, we), indefinite pronouns (eg, it, those), prepositions (eg, for, after, with), negations (eg, not, never), conjunctions (eg, and, but), and quantifiers (eg, many, few). To capture the nuance of language use similarity within each thread, the percentage of total words for each of the 9 types of function words was then calculated for each post and subsequent replies. On the dyadic level, between each user's original posts and each of his or her received replies, the absolute value of the difference was divided by the total for each category. The LSM score can only range from 0 to 1, with scores closest to 1 reflecting high degrees of style matching [43]. The formula used to calculate the LSM is shown in Table 1.

For example, user A made a post and received 3 replies. The personal pronoun LSM for user A's first post was calculated for each of the 3 replies, AR1ppLSM, AR2ppLSM, and AR3ppLSM. Then, the overall personal pronouns LSM score for user A's first thread was calculated by taking the average of the personal pronouns LSM between user A and all of the received replies under the same thread. In the case of user A receiving 3 replies, the overall personal pronoun LSM score for user A is calculated as the mean of the 3 personal pronoun LSM scores.

As there are 9 separate dimensions of function words that make up the overall LSM score, similar calculations were conducted for each user's thread for each function word category. For each user's thread, the 9 separate mean LSM scores for each category were averaged to yield a total LSM score. The same process was applied to each user's thread. The final LSM score for a given user was calculated by taking the average of all threads.

Number of Received Replies

The number of received replies of the forum users were operationalized as the total number of replies one user received from others.

Change in Language Use

In this study, linguistic markers including first-person singular pronouns and words expressing negative emotion were examined. First-person singular pronouns include pronouns such as *I*, *me*, *mine*, and *my* and negative emotions words include words regarding anxiety, anger, and sadness (eg, *hate*, *worthless*, *afraid*, *cry*). For change in first-person singular pronouns, the percentage of first-person singular pronouns in their first 10 messages was subtracted from the percentage of first-person singular pronouns in their last 10 messages. The same process was applied to the calculation of the change in words expressing negative emotion.

Results

Preliminary Analysis Results

The final forum user reply network included 2061 users, and 62,274 replies. On average, each user received 29.23 (SD 44.09) replies and sent out 29.22 (SD 58.20) replies to others. Each user, on average, engaged in interaction with 11 other users (SD 5.00). The average length of their received replies was 155.85 (SD 144.00) words. Descriptive statistics of the key variables are included in Table 2. Theoretical models and results of the hypotheses testing are shown in Figures 4 and 5.

Table 2. Descriptive statistics and zero-order correlation among variables.

Variables	Betweenness	Constraint	Received replies	LSM[a]	Change in first-person singular pronouns	Values, mean (SD)
Betweenness						0.77 (7.57)
r	—[b]	—	—	—	—	
P value	—	—	—	—	—	
Constraint						0.05 (0.03)
r	−0.14	—	—	—	—	
P value	<.01	—	—	—	—	
Received replies						29.23 (44.09)
r	0.59	−0.46	—	—	—	
P value	<.01	<.01	—	—	—	
LSM						0.21 (0.10)
r	0.08	−0.13	0.20	—	—	
P value	<.01	<.01	<.01	—	—	
Change in first-person singular pronouns						0.01 (2.29)
r	−0.01	−0.03	0.01	0.01	—	
P value	—	—	—	—	—	
Change in words expressing negative emotion						0.17 (1.38)
r	−0.10	−0.05	−0.05	−0.01	0.15	
P value	<.01	<.05	<.05	—	<.01	

[a] LSM: language style matching.

[b] Not applicable.

Figure 4. Theoretical framework and results for Hypothesis 1, Hypothesis 2, and Hypothesis 3.

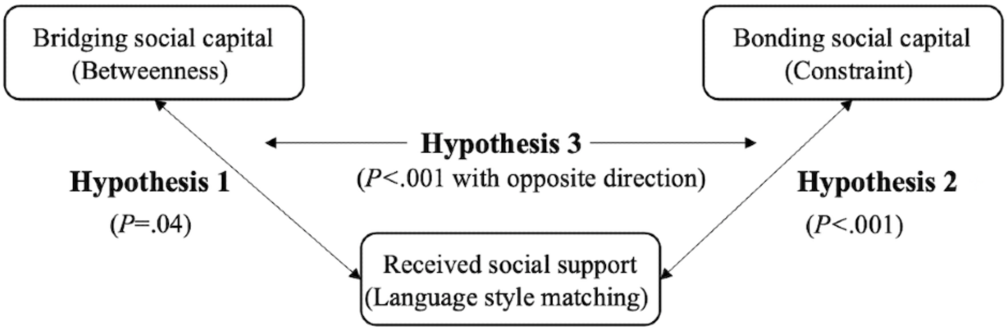
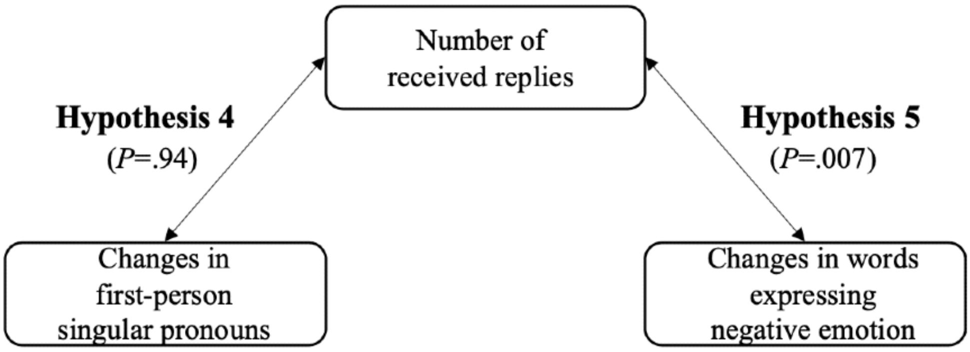


Figure 5. Theoretical framework and results for Hypothesis 4 and Hypothesis 5.



Hypotheses Testing Results

H1 predicted a positive association between users' bridging social capital (betweenness) and their overall LSM. As the LSM score is likely to be affected by the number of users one receives replies from, it was then included in the regression model as a control variable. The analysis revealed that, after controlling for the number of users one received replies from, betweenness (mean 0.77, SD 7.57) was positively associated with the overall LSM (mean 0.21, SD 0.10; β=.06; t_{2059}=2.05; *P*=.04). Bridging social capital explained a small but significant portion of the variance in users' overall LSM (R^2=0.03; $F_{1,2059}$=43.84; *P*=.001). Therefore, H1 was supported.

H2 predicted a positive association between users' bonding social capital (constraint) and their overall LSM. After controlling for the number of users, one received replies from constraint (mean 0.05, SD 0.03) was negatively associated with the overall LSM (mean 0.21, SD 0.10; β=−.17, t_{2059}=−6.72; *P*<.001). Bonding social capital explained a significant portion of the variance in users' overall LSM (R^2=0.04; $F_{1,2059}$=63.24; *P*<.001). This finding was not consistent with the hypothesis. Therefore, H2 was not supported.

H3 predicted that the association between forum users' bonding social capital and LSM would be stronger compared with the association between forum users' bridging social capital and LSM. Standardized *z* scores for LSM were entered as dependent variables, whereas standardized betweenness and constraint scores were entered as independent variables [45]. The corresponding 95% CIs were estimated via bias-corrected bootstrapping (1000 resampled). The CI for betweenness was 0.04 to 0.38, whereas the CI for constraint was −0.19 to −0.05. Therefore, the proportion overlap between the 2 CIs was negative (*P*<.001). This result showed that the association between bonding social capital and LSM was stronger than the association between bridging social capital and LSM. However, the direction of the association between bonding social capital and LSM was the opposite of the prediction. Therefore, H3 was not supported.

H4 predicted a positive association between the users' received replies and their use of first-person singular pronouns change. As the users' language use change is likely to be affected by the number of replies they sent out, it was then included in the regression model as a control variable. Considering that the change in each user's use of first-person singular pronouns was calculated based on the differences between their first and last 10 messages, each user's length of forum use on the forum could also affect their change in language style. Therefore, each user's length of forum use was included in the regression model as a control variable. After controlling for the number of sent-out replies and length of forum use, regression analyses showed that the number of received replies (mean 29.23, SD 44.09) was not significantly associated with the change in their use of first-person singular pronouns (mean 0.01, SD 2.29; β=.01; t_{2059}=0.07; *P*=.94). Therefore, H4 was not supported.

H5 predicted a negative association between the users' number of received replies and their use of words expressing negative emotions changed over time. Similar to the analysis of H4, each user's number of sent-out replies and length of forum use were also included in the regression model as control variables. After controlling for the number of sent-out replies and length of forum use, regression analyses showed that users' number of received replies (mean 29.23, SD 44.09) was negatively associated with the change in their use of words expressing negative emotions changed over time (mean 0.17, SD 1.38; β=−.06; t_{2059}=−2.69; *P*=.007). The users' number of received replies explained a significant portion of the variance in users' use of words expressing negative emotions changed over time (R^2=0.004; $F_{1,2059}$=4.39; *P*=.01). H5 was supported.

Discussion

Principal Findings

Depression, as a common mental disorder, has been estimated to affect more than 264 million people worldwide [46]. Owing to its detrimental effects on the individual's psychological and physical health, scholars and practitioners are dedicated to the prevention, diagnosis, and treatment of depression. Other than seeking help from professionals, individuals with depressive symptoms sometimes turn to online support groups and forums to seek, provide, and exchange social support. Online support groups and forums provide them with access to others who share similar experiences at a low cost. Therefore, studying peer-to-peer support exchanged in these online support groups and forums posits both practical and theoretical importance.

Adopting conceptualizations of social capital and its 2 forms, this study differentiated between 2 distinctive patterns (bridging reflected in betweenness and bonding reflected in constraint) of participation behaviors of online forum users. This study also investigated the associations between online support forum users' social capital and social support in the form of linguistic accommodation. Furthermore, this study also examined the association between receiving replies and health conditions as reflected in forum users' language use. The findings and their implications are discussed in the following paragraphs.

First, the associations between forum users' social capital and their received support were examined. Using social network structural measurement of social capital based on the users' reply patterns, this study found that forum users' bridging social capital was positively associated with the received support, as reflected in the level of linguistic accommodation. This result is in line with previous research examining the association between the positions of online network members and their influences on other networks [9] as well as the association between network positions and their perceived or received support from the network [6]. Network members who possess more central positions tend to stay active in health communities and are more likely to share informational support with less central members [47]. When online forum users engage in interactions with other members, they invest their limited time and attention. From the cost-benefit perspective of social capital, this invested social capital in the form of attending to the support-seeking efforts of others can bring them benefits. For example, the users' bridging capital has been positively associated with the diversity of received replies, whereas users' bonding capital has been positively associated with the length of received replies [6]. Similarly, this study also found that users serve as bridges that connect otherwise unconnected others together may also receive a higher level of linguistic accommodation from others.

However, in terms of the association between users' bonding social capital and received replies, the result indicated an opposite pattern. Forum users' bonding social capital was negatively associated with the level of linguistic accommodation in the received replies. Communication accommodation is a complex phenomenon that can be promoted by various perceptions and social processes. One line of research sees communication accommodation as a reflection of mutual trust, liking, or rapport [48], whereas another line of research treats communication accommodation as being promoted by social perceptions reflecting engagement as well as expertise judgment [49]. Predictions rooted in the rapport perspective would support a positive association between LSM and bonding social capital because the latter facilitates the building of trust and emotional bonds among members, whereas researchers focusing on the engagement perspective see LSM as a reflection of communication partners' involvement and engagement. For example, the use of words expressing emotion from posters can elicit a higher level of engagement and involvement from others [44].

Additional analysis showed that the bonding capital of the forum users was negatively associated with their use of words expressing emotion ($\beta=-.05$; $t_{2059}=-2.34$; $P=.02$). In other words, the more users possessed bonding social capital, the less they will use emotional words in their posts. The less they used emotional words in their posts; the less others were mimicking their language style when replying. The findings on the negative association between bonding social capital and LSM suggest that the underlying mechanism of LSM could be multifaceted, and accommodation could be a function of many social processes such as engagement rather than building rapport or liking.

Expertise judgment, as outlined in expectation states theory (EST), can also shed light on unexpected findings in which a negative association between bonding social capital and LSM was observed [50]. According to EST, individuals use characteristics indicating pre-existing status (eg, gender, race) and behavioral interchange patterns within-group processes to judge the competence of others. Characteristics indicating pre-existing status can be directly communicated or inferred through speech and nonverbal cues, whereas interchange patterns can be observed during interactions with others [49]. In this study, the language used by a given user and the linguistic accommodation of others can serve as characteristics indicating pre-existing status and behavioral interchange patterns, respectively. For users who possess more bonding social capital, their language style is less affective and less accommodated by others. They can both serve as grounds for others to form expert judgments. Therefore, users who possess more bonding capital are likely to be viewed as possessing less expertise and a low social status. Thus, others are less likely to accommodate their language style. Although without further examination of the users' motives or subjective experience of their interactions with other forum users, it remains an empirical question as to what cues forum users utilize to form judgments or perceptions about other users. However, in text-based online support forums with a limited number of cue systems, the language use of forum users can nonetheless serve as an important ground for users to form perceptions about each other.

One caveat regarding the association between the users' social capital and LSM in their received replies lies in the direction of the association. Social capital and social support theories may evaluate LSM as an outcome or reflection of interpersonal and social processes, whereas virtual team dynamics and related computer-mediated communication (CMC) theories may see network positions and formations as outcomes of LSM because of the text-based nature of CMC. For example, language features of the user him or herself and the perceived LSM of the user can both be associated with the user's network position. As the LSM was calculated between a user's original post and the received replies to the same post, language features of the poster could affect the level of linguistic adaptation from others. The forum user's own language features can also be affected by their network positions or the composition of their social networks. For example, the heterogeneity of the social networks of Facebook users is negatively associated with language style variability, as reflected in their status updates [34]. The lowest common denominator approach to online self-presentation posits that individuals will choose to share information that is considered appropriate for all audience segments [51]. The proposed pattern has been observed in the status updates of Facebook users: users tend to vary little in their language style when the audience segments are disconnected with each other [34]. Online support forum users who possess more bridging social capital (higher network betweenness) share ties with heterogeneous groups of others. On the basis of the lowest common denominator approach, these users are assumed to have less variability in their language style. Therefore, the replies they received from others share more similarities with their own language style, given that a user's LSM is calculated between each post contributed by the same user and each reply addressing the same post.

The results of this study showed a positive association between the number of received replies and a decrease in the use of words expressing negative emotion. These results are in line with the self-awareness theory of depression [52] and previous literature on language features of depressed people. As discussed earlier, words expressing negative emotion are important linguistic markers for identifying depression. Depressed individuals are more self-focused and express more negative emotions [30]. Previous studies have established that language features can be markers of mental health [32]. The results suggested that for depression forum users, after controlling for their length of forum use, receiving replies was associated with improvement in their psychological states, although causality could not be established because of the nature of the data. Further mediation analysis showed that the level of communication accommodation exemplified in the users' received replies did not mediate the effects of bridging and bonding social capital on the users' change in their use of self-referent words or words expressing negative emotion (Multimedia Appendix 3 shows mediation results).

Limitations

Several limitations of the study are worth noting. The first limitation lies in the lack of causality in the findings. As the users' bridging and bonding social capital, as well as the LSM, was treated cross-sectional as individual attributes, causal relations cannot be established. Although the changes in the language features of forum users can serve as longitudinal indicators of their language use tendency, the data were treated as cross-sectional. Although the benefit of web-scraped forum users' participation patterns is to offer an unobtrusive way to collect the users' actual behavioral data, future studies could split data into different periods and observe the trend in the users' social capital and language use to establish causal relations between forum participation and related outcome measurements.

The second limitation lies in the lack of content measures for exchanged messages. Although computerized text analyses offer powerful and effective ways to investigate the forum users' language features, the context is largely ignored. Due to the exclusive reliance on words recorded in the analytical software's dictionaries, the assessment of the outcomes of interest can be somewhat rigid. Future research could use human coders to cross-validate computerized text analysis results.

Conclusions

This study contributes to our understanding of how people interact with each other in online support forums. In particular, we extend the research focus of online supportive communication to the network positions of users and the communication accommodation of their received support. Using a social network analysis method to analyze web-scraped data from an online support forum, this study offers an objective and nuanced way to examine the structural aspect of social capital. Analyzing the linguistic features of messaged exchanged in online support forums, this study also offers an unobtrusive way to analyze the actual messages created and used by online support forum users.

Acknowledgments

This research was supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China, grant no. 19XNF027.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Python code for web crawler.
[ZIP File (Zip Archive), 1275 KB - jmir_v22i6e17365_app1.zip]

Multimedia Appendix 2
Python code to process network data.
[ZIP File (Zip Archive), 20 KB - jmir_v22i6e17365_app2.zip]

Multimedia Appendix 3
Mediation results of communication accommodation mediating the effects of 2 types of social capital on users' change in self-referent words and words expressing negative emotion.
[DOCX File , 13 KB - jmir_v22i6e17365_app3.docx]

References

1. Bargh JA, McKenna KY. The internet and social life. Annu Rev Psychol 2004;55:573-590. [doi: 10.1146/annurev.psych.55.090902.141922] [Medline: 14744227]
2. Adler P, Kwon S. Social capital: prospects for a new concept. Acad Manag Rev 2002 Jan;27(1):17-40. [doi: 10.5465/amr.2002.5922314]
3. Ellison NB, Steinfield C, Lampe C. The benefits of Facebook 'friends:' social capital and college students' use of online social network sites. J Comput Commun 2007;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
4. High A, Solomon D. Locating computer-mediated social support within online communication environments. In: Wright K, Webb L, editors. Computer-Mediated Communication in Personal Relationships. New York, USA: Peter Lang Publishing; 2011:119-155.
5. Wright K, Rains S, Banas J. Weak-tie support network preference and perceived life stress among participants in health-related, computer-mediated support groups. J Comput Commun 2010;15(4):606-624. [doi: 10.1111/j.1083-6101.2009.01505.x]
6. Pan W, Shen C, Feng B. You get what you give: understanding reply reciprocity and social capital in online health support forums. J Health Commun 2017 Jan;22(1):45-52. [doi: 10.1080/10810730.2016.1250845] [Medline: 28027009]

7. Chung JE. Social networking in online support groups for health: how online social networking benefits patients. J Health Commun 2014;19(6):639-659. [doi: 10.1080/10810730.2012.757396] [Medline: 23557148]
8. Himelboim I. Reply distribution in online discussions: a comparative network analysis of political and health newsgroups. J Comput Commun 2008;14(1):156-177. [doi: 10.1111/j.1083-6101.2008.01435.x]
9. Huffaker D. Dimensions of leadership and social influence in online communities. Hum Commun Res 2010;36(4):593-617. [doi: 10.1111/j.1468-2958.2010.01390.x]
10. Drentea P, Moren-Cross JL. Social capital and social support on the web: the case of an internet mother site. Sociol Health Illn 2005 Nov;27(7):920-943 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00464.x] [Medline: 16313523]
11. Resnick P. Beyond bowling together: sociotechnical capital. In: Carroll J, editor. Human-Computer Interaction in the New Millennium. Boston, UK: Addison-Wesley; 2001:247-272.
12. Putnam R. Bowling Alone: The Collapse and Revival of American Community. New York, USA: Simon and Schuster; 2001.
13. Granovetter M. The strength of weak ties. Am J Sociol 1973 May;78(6):1360-1380. [doi: 10.1086/225469]
14. Burt R. Structural holes and good ideas. Am J Sociol 2004 Sep;110(2):349-399. [doi: 10.1086/421787]
15. Shen C, Monge P, Williams D. Virtual brokerage and closure. Commun Res 2012 Aug 3;41(4):459-480. [doi: 10.1177/0093650212455197]
16. Poortinga W. Social relations or social capital? Individual and community health effects of bonding social capital. Soc Sci Med 2006 Jul;63(1):255-270. [doi: 10.1016/j.socscimed.2005.11.039] [Medline: 16427171]
17. Toma C. Towards conceptual convergence: an examination of interpersonal adaptation. Commun Q 2014 Apr 9;62(2):155-178. [doi: 10.1080/01463373.2014.890116]
18. Gallois C, Ogay T, Giles H. Communication accommodation theory: a look back and a look ahead. In: Gudykunst W, editor. Theorizing About Intercultural Communication. Thousand Oaks, CA: Sage Publications; 2005:121-148.
19. Rains S. Language style matching as a predictor of perceived social support in computer-mediated interaction among individuals coping with illness. Communic Res 2015 Jan 13;43(5):694-712. [doi: 10.1177/0093650214565920]
20. Wright K. Communication in health-related online social support groups / communities: a review of research on predictors of participation , applications of social support theory , and health outcomes. Rev Commun Res 2016;4:65-87. [doi: 10.12840/issn.2255-4165.2016.04.01.010]
21. Niederhoffer K, Pennebaker J. Linguistic style matching in social interaction. J Lang Soc Psychol 2016 Jul 25;21(4):337-360. [doi: 10.1177/026192702237953]
22. Rains S, Young V. A meta-analysis of research on formal computer-mediated support groups: examining group characteristics and health outcomes. Human Commun Res 2009;35(1):309-336. [doi: 10.1111/j.1468-2958.2009.01353.x]
23. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
24. Wright K. Perceptions of on ‐ line support providers: an examination of perceived homophily, source credibility, communication and social support within on ‐ line support groups. Commun Q 2000 Jan;48(1):44-59. [doi: 10.1080/01463370009385579]
25. Eastin MS, LaRose R. Alt.support: modeling social support online. Comput Human Behav 2005 Nov;21(6):977-992. [doi: 10.1016/j.chb.2004.02.024]
26. Albrecht T, Goldsmith D. Social support, social networks, and health. In: Thompson T, Dorsey A, Miller K, Parrott R, editors. The Routledge Handbook of Health Communication. New Jersey, USA: Lawrence Erlbaum Associates Publishers; 2003:263-284.
27. Wright KB, Bell SB, Wright KB, Bell SB. Health-related support groups on the internet: linking empirical findings to social support and computer-mediated communication theory. J Health Psychol 2003 Jan;8(1):39-54. [doi: 10.1177/1359105303008001429] [Medline: 22113899]
28. Cummings JN, Sproull L, Kiesler SB. Beyond hearing: where the real-world and online support meet. Gr Dyn Theory, Res Pract 2002;6(1):78-88. [doi: 10.1037//1089-2699.6.1.78]
29. Chung C, Pennebaker J. The psychological functions of function words. In: Fiedler K, editor. Social Communication. London, UK: Psychology Press; 2007:343-359.
30. Tausczik Y, Pennebaker J. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 8;29(1):24-54. [doi: 10.1177/0261927x09351676]
31. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 2001;63(4):517-522. [doi: 10.1097/00006842-200107000-00001] [Medline: 11485104]
32. Rude S, Gortner E, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
33. Road to Recovery - Depression Forums Depression Forums. URL: https://www.depressionforums.org/forums/ [accessed 2020-05-11]
34. Gil-Lopez T, Shen C, Benefield G, Palomares N, Kosinski M, Stillwell D. One size fits all: context collapse, self-presentation strategies and language styles on Facebook. J Comput Commun 2018;23(3):127-145. [doi: 10.1093/jcmc/zmy006]

35. Kosinski M, Bachrach Y, Kohli P, Stillwell D, Graepel T. Manifestations of user personality in website choice and behaviour on online social networks. Mach Learn 2013 Oct 19;95(3):357-380. [doi: 10.1007/s10994-013-5415-y]
36. Butts C. Comprehensive R Archive Network (CRAN). 2019. Tools for Social Network Analysis R package SNA Version 2 URL: https://CRAN.R-project.org/package=sna [accessed 2020-03-18]
37. Butts C. Comprehensive R Archive Network (CRAN). 2019. Classes for Relational Data R Package Network Version 1 URL: https://CRAN.R-project.org/package=network [accessed 2020-03-18]
38. Pennebaker J, Boyd R, Jordan K, Blackburn K. The University of Texas at Austin. 2015. The Development and Psychometric Properties of LIWC2015 URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2020-03-18]
39. Ganley D, Lampe C. The ties that bind: social network principles in online communities. Decis Support Syst 2009 Jun;47(3):266-274. [doi: 10.1016/j.dss.2009.02.013]
40. Hanneman R, Riddle M. University of California. 2005. Introduction to Social Network Methods URL: https://faculty.ucr.edu/~hanneman/nettext/ [accessed 2020-05-15]
41. Newman MJ. A measure of betweenness centrality based on random walks. Soc Netw 2005 Jan;27(1):39-54. [doi: 10.1016/j.socnet.2004.11.009]
42. Burt R. Brokerage and Closure: an Introduction to Social Capital. Oxford, UK: Oxford University Press; 2005.
43. Gonzales A, Hancock J, Pennebaker J. Language style matching as a predictor of social dynamics in small groups. Commun Res 2009 Nov 4;37(1):3-19. [doi: 10.1177/0093650209351468]
44. Ireland ME, Pennebaker JW. Language style matching in writing: synchrony in essays, correspondence, and poetry. J Pers Soc Psychol 2010 Sep;99(3):549-571. [doi: 10.1037/a0020386] [Medline: 20804263]
45. Cumming G. Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. New York, USA: Routledge; 2013.
46. World Health Organization. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates URL: https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/ [accessed 2020-04-28]
47. Introne J, Semaan B, Goggins S. A Sociotechnical Mechanism for Online Support Provision. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7-12, 2016; San Jose, USA p. 3559-3571. [doi: 10.1145/2858036.2858582]
48. Kozlowski S, Ilgen D. Enhancing the effectiveness of work groups and teams. Psychol Sci Public Interest 2006 Dec;7(3):77-124. [doi: 10.1111/j.1529-1006.2006.00030.x] [Medline: 26158912]
49. Liao W, Bazarova N, Yuan Y. Expertise judgment and communication accommodation in linguistic styles in computer-mediated and face-to-face groups. Commun Res 2016 Feb 5;45(8):1122-1145. [doi: 10.1177/0093650215626974]
50. Wagmer D, Berger J. Expectation states theory: an evolving research program. In: Berger J, Zelditch Jr M, editors. New Directions in Contemporary Sociological Theory. New York, USA: Rowman & Littlefield; 1982:41-76.
51. Hogan B. The presentation of self in the age of social media: distinguishing performances and exhibitions online. Bull Sci Technol Soc 2010 Nov;30(6):377-386. [doi: 10.1177/0270467610385893]
52. Pyszczynski T, Greenberg J. Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression. Psychol Bull 1987;102(1):122-138. [doi: 10.1037/0033-2909.102.1.122]

Abbreviations

CMC: computer-mediated communication
EST: expectation states theory
JSON: JavaScript Object Notation
LIWC: Linguistic Inquiry and Word Count

Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by J Meng, S Kumar, G Khalil; comments to author 10.03.20; revised version received 21.04.20; accepted 27.04.20; published 24.06.20.

Please cite as:
Pan W, Feng B, Shen C
Examining Social Capital, Social Support, and Language Use in an Online Depression Forum: Social Network and Content Analysis
J Med Internet Res 2020;22(6):e17365
URL: https://www.jmir.org/2020/6/e17365
doi:10.2196/17365
PMID:32579125

Original Paper

Usability of a Consumer Health Informatics Tool Following Completion of a Clinical Trial: Focus Group Study

Hwayoung Cho[1], PhD; Tiffany Porras[2], MPH; Gabriella Flynn[2], MEd; Rebecca Schnall[2], PhD, MPH

[1]College of Nursing, University of Florida, Gainesville, FL, United States
[2]School of Nursing, Columbia University, New York, NY, United States

Corresponding Author:
Rebecca Schnall, PhD, MPH
School of Nursing
Columbia University
560 West 168th Street
New York, NY, 10032
United States
Phone: 1 2123426886
Email: rb897@columbia.edu

Abstract

Background: Mobile health (mHealth) apps have the potential to be effective tools for encouraging patients with chronic diseases to self-manage their health. The success of mHealth apps is related to technology acceptance and its subsequent use by intended consumers. Therefore, it is essential to gain insights from consumers' perspectives about their use of mHealth apps in daily life.

Objective: The purpose of this work was to understand consumers' perspectives on use of a self-management app following completion of a clinical trial that tested the efficacy of the app for improving health outcomes.

Methods: We conducted five focus groups with paricipants of a clinical trial (NCT03182738) who were randomized to use the video information provider (VIP) for HIV-associated nonAIDS (HANA) conditions app (VIP-HANA) or an attention control app. Thematic analysis was conducted, and the themes were organized according to the two key constructs of the technology acceptance model framework: perceived usefulness and perceived ease of use.

Results: Thirty-nine people living with HIV (20 from the intervention group and 19 from the control group) participated in the focus group sessions. Of the eight themes identified from focus group data, the five themes related to perceived usefulness were: (1) self-monitoring HIV-related symptoms of HANA conditions, (2) enhanced relationship with clinical providers, (3) improvement in physical and emotional health, (4) long-term impact of self-care strategies on improvement in symptoms of HANA conditions, and (5) inspired lifestyle changes to manage symptoms. The three themes related to perceived ease of use were: (1) easy to navigate, (2) avatar personalization, and (3) privacy/confidentiality maintained even when changing the location of app use.

Conclusions: Perceived ease of use was similar in both study groups but perceived usefulness differed between study groups. Participants in both study groups found the VIP-HANA app to be useful in monitoring their symptoms and enhancing communication with their clinical care providers. However, only intervention group participants perceived the app to be useful in improving overall health and long-term symptom management. Findings from this study highlight factors that are essential to ensure the usefulness of self-management apps and facilitate sustained use of mHealth apps for people living with chronic illnesses.

(J Med Internet Res 2020;22(6):e17708) doi:10.2196/17708

KEYWORDS

consumer health informatics tool; mobile Health; mobile apps; clinical trial; symptom care; self-management; HIV-associated nonAIDS (HANA); HANA conditions; HIV/AIDS

Introduction

With the widespread proliferation of smartphones, the use of mobile technology has rapidly increased in health care [1,2]. Consumer health informatics tools such as mobile health (mHealth) apps have the potential to be effective tools to encourage patients with chronic diseases to self-manage their health [3-7]. The video information provider (VIP) for HIV-associated nonAIDS (HANA) conditions app (VIP-HANA) is an mHealth app that was developed to help people living with HIV self-manage their symptoms by providing self-care strategies [8]. Development of the VIP-HANA app is described elsewhere [9]. In short, the VIP-HANA app is comprised of 728 self-care strategies for the following 28 HANA condition-related symptoms: fatigue, depression, muscle aches, difficulty falling asleep, anxiety, difficulty staying asleep, difficulty remembering, neuropathy, difficulty concentrating, decreased sex drive, diarrhea, maintaining an erection, shortness of breath, constipation, dry mouth, clumsiness, weight loss, dizziness, heartburn, dry eyes, changes in appetite, ringing in ear, cough, nausea/vomiting, fever, difficulty with urination, speech difficulties, and pain during sex [10]. The 728 self-care strategies were tailored based on the symptom, gender, race/ethnicity, and HANA condition of each user. An avatar guides users through the app. Patients then receive self-care strategies with animated videos for ameliorating their symptoms. Users review their symptom experience over time with symptom reports, set a weekly email reminder to complete sessions in the app, and email/download the history of their reported symptoms and suggested self-care strategies [8].

Two versions of the app were created to assess the efficacy of the VIP-HANA app in improving the symptom burden in a 6-month randomized controlled trial (RCT) (NCT03182738) that included 100 people living with HIV with HANA conditions (50 randomly assigned to the intervention group, 50 assigned to the control group). All participants were asked to log in at least once per week. Both study groups received weekly symptom assessment questions through the VIP-HANA app, yet only the intervention group participants were provided with the self-care strategies. Multimedia Appendix 1 presents a comparison of the two versions of the VIP-HANA app features. Participants in both groups were given US $5 per week for their time and data costs associated with completing the weekly symptom assessments.

According to a recent US national survey of mHealth app use, approximately half (46%) of the respondents reported downloading a health app but were no longer using it or had uninstalled it [11]. Given that the success of mHealth apps is related to the acceptance of technology and its subsequent use by intended consumers, there is a need to gain in-depth insights from consumers' perspectives about use of mHealth apps in their daily life. Accordingly, the purpose of this study was to understand consumers' perspectives on mHealth use (VIP-HANA app) following an RCT of an HIV symptom self-management intervention for people living with HANA conditions.

Methods

Sampling and Recruitment

This study was a follow up of a 6-month trial including 100 people living with HIV with HANA conditions (50 randomly assigned to the intervention group, 50 to the control group). We used a convenience sampling approach to recruit a subsample of our study participants to participate in one of five focus group sessions followed by completion of the study trial; the focus group sessions took place once per month [12-14].

Procedures

The Institutional Review Board of Columbia University Medical Center reviewed and approved all research activities. We conducted three focus group sessions with intervention group participants (N=20) and two focus groups sessions with control group participants (N=19). Prior to the focus group sessions, all participants were provided an explanation of the study procedures and completed a signed consent form. Using open-ended questions via a semistructured focus group guide designed based on the app's features/functionality (Textbox 1), a moderator (a faculty member at Columbia University School of Nursing) facilitated the focus group sessions, and participants were encouraged to discuss issues regarding their perceptions related to use of the VIP-HANA app (intervention/control group versions). All focus group sessions were audio-recorded. Data collection continued until saturation of themes was reached. Participants were compensated with US $30 for their time.

Textbox 1. Focus group guide based on 6-month experience of using the video information provider for HIV-associated nonAIDS conditions (VIP-HANA) app.

Experience Using VIP-HANA

- Please describe your experience using the VIP-HANA app in social settings.
    - Probe: where and when did you use it most (house/workplace/clinic/café; after breakfast/when commuting, etc) and why?
- How did this fit into your lifestyle and schedule?
- Please describe your experience navigating the app pages.
    - Probe: basic structure of the app
- After your first app use, how easy/difficult was it for you in follow-up uses?
- What are your thoughts about the design of the VIP-HANA app?
    - Probe: main logo/avatars; font/color; progress bar; Remember Me (save ID/password); Log-In Help; Continue/Skip button
- Please describe your experience with any technical issues.
    - Probe: Log-In, crash, Back button, Continue/Skip button, How Our App Works, reminder/email, error message

(Intervention group only)

- What are your thoughts about the videos displayed in the VIP-HANA app?
    - Probe: watching videos vs reading the content; videos with sounds, the type of sounds; effect of data plan on watching the videos
- What was your experience using the app to review your symptoms and self-care strategies (=Your History)?

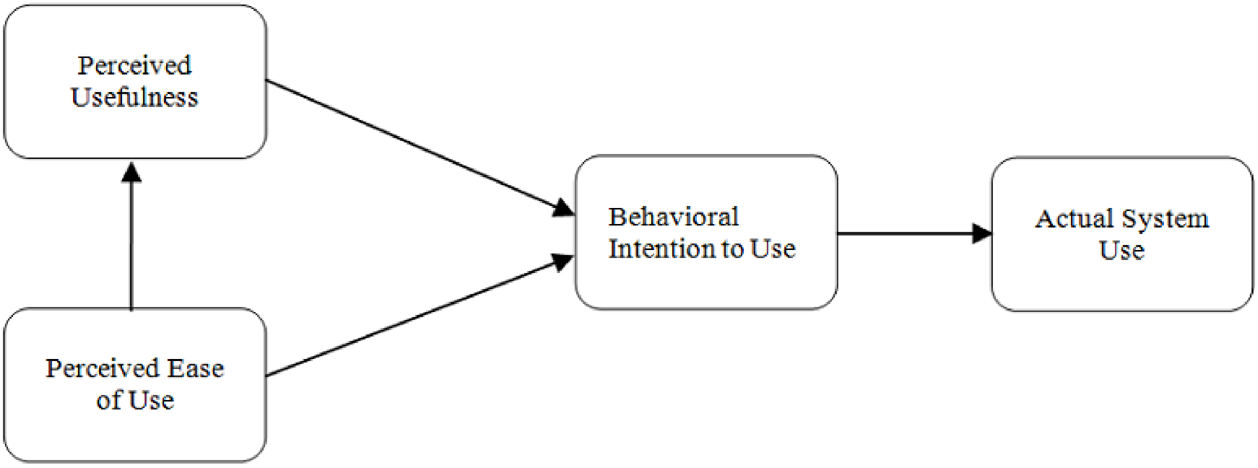
Impact of VIP-HANA

- How did this app help you gain information about relieving your symptoms?
- How do you think that your symptoms changed after using the VIP-HANA app for 6 months?
- How confident are you in your ability to self-manage your symptoms?
- What are some of the ways that your health might benefit through the use of VIP-HANA?
    - Probe: changes of current personal, professional, or health care provider relationships
- Please describe how the app did/can change your quality of life.

Data Analysis

The technology acceptance model (TAM) was used as a guide for the data analysis [15]. The TAM includes the following two key constructs: perceived usefulness and perceived ease of use. Perceived usefulness represents the degree to which an individual believes that using a particular system would enhance one's job performance, whereas perceived ease of use signifies the degree to which an individual believes that using a particular system would be free of effort. The TAM suggests that these two constructs of user acceptance influence the behavioral intention to use a system and predict the extent of adoption of the system (Figure 1).

Figure 1. Technology acceptance model (TAM).



Audio recordings from the focus groups were transcribed. Thematic analysis was used to explore significant themes with similar patterns across focus group sessions. Research team members (HC, TP, and GF) independently reviewed the transcripts at least twice, and two research team members (TP and GF) generated a set of codes through a combination of inductive and deductive content analysis [16-18]. A codebook was developed using a Microsoft Excel spreadsheet, and free text excerpted from the transcripts was entered into the codebook followed by each of the themes. Coding discordance between TP and GF was reconciled by a third researcher (HC). Guided by the TAM framework (Figure 1) [15], each theme in the codebook was categorized into the two constructs of perceived usefulness and perceived ease of use.

Results

Sample

A total of 39 people who completed the 6-month trial participated in the five focus group sessions. The mean age of the study participants was 55 years (SD 10.70). Additional characteristics of the study sample are presented in Table 1.

Table 1. Characteristics of the study participants (N=39).

Characteristic	All participants (N=39)	Intervention (N=20)	Control (N=19)
Gender, n (%)			
Male	22 (56)	12 (60)	10 (53)
Female	15 (38)	7 (35)	8 (42)
Transgender	1 (3)	0 (0)	1 (5)
Genderqueer	1 (3)	1 (5)	0 (0)
Sex at birth, n (%)			
Male	24 (62)	13 (65)	11 (58)
Female	15 (38)	7 (35)	8 (42)
Race, n (%)			
African American/Black	27 (69)	13 (65)	14 (74)
Other	8 (21)	4 (20)	4 (21)
White	4 (10)	3 (15)	1 (5)
Hispanic/Latino ethnicity, n (%)	13 (33)	6 (30)	7 (37)
Marital status, n (%)			
Single	22 (56)	12 (60)	10 (53)
In a relationship	13 (33)	6 (30)	7 (37)
Legally married/domestic partnership	2 (5)	1 (5)	1 (5)
Widowed	1 (3)	0 (0)	1 (5)
Divorced	1 (3)	1 (5)	0 (0)
Education level, n (%)			
High school diploma or equivalent	14 (36)	6 (30)	8 (42)
Some college	10 (26)	6 (30)	4 (21)
Some high school, no diploma	6 (15)	3 (15)	3 (16)
Associate or technical degree	4 (10)	1 (5)	3 (16)
Bachelor/college degree	2 (5)	1 (5)	1 (5)
Professional or graduate degree	2 (5)	2 (10)	0 (0)
None	1 (3)	1 (5)	0(0)
Current employment status, n (%)			
Disabled	16 (41)	8 (40)	8 (42)
Unemployed	12 (31)	6 (30)	6 (31)
Retired	5 (13)	3 (15)	2 (11)
Working part time	4 (10)	3 (15)	1 (5)
Retired and other	2 (5)	0 (0)	2 (11)
Working full time	0 (0)	0 (0)	0 (0)
Annual income, n (%)			
Less than US $10,000	18 (46)	11 (55)	7 (37)
US $10,000-$19,000	8 (21)	3 (15)	5 (26)
US $20,000-$39,999	7 (17)	4 (20)	3 (16)
Don’t know	5 (13)	1 (5)	4 (21)
US $40,000-$59,000	1 (3)	1 (5)	0 (0)
Health insurance provider, n (%)			
Public (Medicare, Medicaid, Ryan White)	32 (82)	16 (80)	16 (84)

Characteristic	All participants (N=39)	Intervention (N=20)	Control (N=19)
Other (eg, ADAP[a], Veteran's Association, multiple)	6 (15)	3 (15)	3 (16)
None	1 (3)	1 (5)	0 (0)
Private (through employer)	0 (0)	0 (0)	0 (0)

[a] ADAP: AIDS drug assistance program.

Main Themes

A total of eight themes related to use of the VIP-HANA app were identified. The themes were organized by the two key constructs of the TAM framework [15], including five themes related to perceived usefulness and three themes related to perceived ease of use. Two themes to enhance the acceptance of the app were additionally identified.

Perceived Usefulness

The five themes related to perceived usefulness were: (1) self-monitoring HIV-related symptoms of HANA conditions, (2) enhanced relationship with clinical providers, (3) improvement in physical and emotional health, (4) long-term impact of self-care strategies on improvement in symptoms of HANA conditions, and (5) inspired lifestyle changes to manage symptoms.

Self-Monitoring HIV-Related Symptoms of HANA Conditions

Participants in both the intervention and control groups felt that the app was helpful for improving self-monitoring of their symptoms. A control group participant expressed that the app helped them realize the frequency of certain symptoms they were having, by stating:

> *I found that it really kept me in touch with my symptoms. And a lot of things were happening that I really didn't realize until I started using the app.* [FG2]

An intervention group participant mentioned:

> *The app shows you certain things that you can have, and what you can do. And you work with that. And it's been great for me. I was getting more confident with tracking and managing my symptoms.* [FG3]

Moreover, intervention group participants described enjoyment in tracking their HANA conditions:

> *So for those of us that do have secondary illnesses such as diabetes, asthma, high blood pressure, and so forth and so on, the app was a great thing, and not just helping us navigate the system and our bodies as far as HIV is concerned, but helped us in those other issues as well.* [FG1]
>
> *And she said the app is wonderful… for diabetes, hepatitis C, and for the different harm reduction. We know what symptoms I have.* [FG3]

Enhanced Relationship With Clinical Providers

Participants in both the intervention and control groups stated that they showed the app to their doctors/clinical therapists to guide their conversations and share their symptom reports, which empowered them to better communicate with their clinical providers and enrich their relationship. An intervention group participant described:

> *The changes that I made were, the app was very useful in that, in conjunction with my doctor. Because I also showed this to my doctor… And they all agreed that it's a wonderful thing.* [FG1]

A control group participant stated:

> *I liked the app because it was times I wanted to ask my doctor questions, but I didn't. But by doing this here app, it made me take a pen and paper and write down what I want to say to her next time I see her.* [FG2]

Improvement in Physical and Emotional Health

Participants in the intervention group noted improvements in their health as a result of app use. Participants commented that the app helped them work through their emotional issues. An intervention group participant stated:

> *It is not just HIV health. It is diabetes, asthma, high blood pressure… and so on… Not just your physical and your medications and stuff like that. But things that you can do to make yourself happier, better, more social.* [FG1]

Long-Term Impact of Self-Care Strategies on Improvement in Symptoms of HANA Conditions

Intervention group participants stated that they continued to use the self-care strategies for their symptoms even after the study was over. They considered that the self-care strategies resulted in a lasting improvement of their symptoms and therefore they continued to use them. An intervention group participant mentioned:

> *When I go through stress, I get those migraine headaches. In the app, it tells you to relax, turn off the TV. Turn off everything and just be calm in a dark room. And that still helps me. I am still doing that.* [FG1]

Inspired Lifestyle Changes to Manage Symptoms

Participants in the control group noted that the symptom questions on the VIP-HANA app helped them to think about their recent symptoms and inspired lifestyle changes. For example, some control group participants stated:

> *I remember one time, I was really depressed that week. And I put down real severe depression. And it made me think... So I'm trying to figure out why am I depressed? What made me depressed this week? What should I do? What changes?* [FG2]

> *But when you're looking at, when you're pressing this button several times every, several times a month, with the same thing, it reminds you. You need to really get serious about what you've been thinking about all this time.* [FG2]

Perceived Ease of Use

The three major themes related to perceived ease of use were: (1) easy to navigate, (2) avatar personalization, and (3) privacy/confidentiality maintained even when changing the location of app use.

Easy to Navigate

Participants in both the intervention and control groups found that the VIP-HANA app was easy to navigate, and they were able to use the app without assistance and did not experience technical issues. An intervention group participant stated:

> *Nothing hard at all. I am not very good to the tech… but this app, easy to get to start… easy to go to the next… very easy to move forward.* [FG3]

A control group participant described:

> *Yeah. The app I thought was pretty easy to navigate through. And the questions were basically always the same. I didn't need any help although I am not good at technology.* [FG2]

Avatar Personalization

Participants in the intervention and control groups noted the usefulness of an avatar to guide them through the VIP-HANA app and help them navigate through the app. They also noted that the avatar promoted their engagement with the app. Some participants in the intervention group mentioned:

> *I feel like I connected to it more by choosing the one that best looked like me or was fitted to me.* [FG1]
>
> *When you see it in the app, where you're actually seeing the cartoon character [avatar], get up and going. It encourages you that you can get up and go. I think it was very practical.* [FG5]

Some participants in the control group stated:

> *I think when you look at it [avatar], it's more personal. It's about you. And I think that it's good for your psyche to realize that you've got to really take care of yourself.* [FG2]
>
> *The app I thought was pretty easy to navigate through because of the avatar. The avatar looks like me… myself!* [FG4]
>
> *I was going to mention that the whole thing is laid out pretty attractively. The little avatar made it, to me, it made it almost like a game.* [FG2]

Privacy/Confidentiality Maintained Even When Changing Location of App Use

Participants in both the intervention and control groups felt comfortable using the app in a variety of locations ranging from their home to doctor's offices and in public spaces (ie, everywhere). They thought that the app was very discrete and did not disclose their HIV status to others if they could see the app on the screen. Several participants in the intervention group described this aspect:

> *To me, I feel like it was enough discretion in the app that if I wanted to explain it to somebody, because I was okay pulling out anywhere… everywhere…* [FG1]
>
> *I feel that that was really good, in the sense of you can do it really anywhere.* [FG5]

Some participants in the control group stated:

> *One of the good things about the program, it doesn't mention anything related to HIV. The reason… If someone want to get into your phone, nobody can access personal information related to the app.* [FG2]
>
> *I was happy because I can use it anytime and anywhere; at work, on the train.* [FG4]

Additional Comments to Enhance App Acceptance

Need for Updates of Self-Care Strategies and Videos

Participants in the intervention group expressed their desire for the app to include more self-care strategies and realistic videos. Participants felt that the app questions (symptoms) and suggestions (self-care strategies) were particularly useful for people living with HIV who were newly diagnosed with HIV/AIDS. They thought that the animated videos showing the suggestions were too simple or not serious enough in some cases. Some participants in the intervention group described:

> *When I look at the study you have to say that it's really geared more to like an HIV 101 plan where it's just basic stuff… For a person that's newly diagnosed and don't know a lot of things about HIV, I think it's the best thing for them.* [FG1]
>
> *Some of it [animated videos] was unrealistic. If you don't have the HIV or the pain or things that we actually experience as patients, you really couldn't understand how serious sometimes it is. It felt to me like they downplayed our experiences with those videos.* [FG3]

Request for Additional Functionality of Symptom Reports

Participants in the control group wanted the app to have an additional functionality to track their symptom reports so that they do not have to remember their weekly answers to the symptom questions (eg, what symptoms they had), as the app feature of symptom reports was limited for the intervention group participants only in this study. For example, a control group participant stated:

> *I don't remember there being an ability to be able to say, to run a report, so to speak, and say; between January and March, I had dry eyes this many times during the month... if I can print something out or if I can open it on the phone and show it to the doctor like that, that would be a nice addition to what you already have.* [FG2]

Discussion

Principal Findings

With the exponential increase of mobile technology use in health care, mHealth apps can be a promising consumer health informatics tool for health behavior change and self-management. Although existing evidence supports the use of mHealth apps to improve health outcomes in people with chronic diseases [3-7], there has been limited assessment of consumers' experience following their use of these apps during a clinical trial. Given the current landscape of discontinuation of mHealth apps use [11], it is imperative to understand how consumers are interacting with the apps in their daily life in an effort to accelerate acceptance of the apps by the consumers as well as ensure their subsequent use. To better understand consumers' use of a self-management app, we included both intervention and control group participants following completion of an RCT that tested the efficacy of the app for improving health outcomes. We captured valuable perspectives of both study group participants on their app use using focus groups guided by a theory-driven framework. This approach facilitated gaining an in-depth understanding of the key factors that would have a critical influence on adoption of the app [15], which is a significant strength of our study.

Participants in both the intervention and control groups perceived the VIP-HANA app to be useful for monitoring their symptoms, and the app enhanced their relationship with their clinical providers. Although self-care strategies of existing symptoms were not provided to the control group participants, all participants described the need to record, track, and share their symptom status with their providers. Participants described that use of the symptom reports would expedite opportunities of health communication and enhance their relationship with their clinical providers. Current research supports that interpersonal communication between HIV patients and their health care provider improves medication adherence and long-term care in HIV, highlighting the importance of patient-provider communication for ensuring that patients remain in the continuum of HIV care [19-21].

Although all three themes related to perceived ease of use were identified in both study groups, only the intervention group participants described the VIP-HANA app as being useful for enhancing their physical and emotional health and supporting long-term symptom management. Existing evidence suggests that perceived usefulness is a stronger factor compared with perceived ease of use for influencing technology adoption, and perceived usefulness has a positive influence on overall use of an app [22,23].

mHealth apps should be visually appealing to consumers [24], but more fundamentally should contain evidence-based health information that consumers can use to improve their health. Participants considered that the app could help improve their health and manage symptoms through the use of self-care strategies. The VIP-HANA app was developed by employing evidence from patient-centered outcomes research studies [25], and the HIV symptom self-care strategies provided within the app were tailored based on symptoms, gender, race/ethnicity, and HANA condition [26,27]. To our knowledge, no mHealth intervention offers health information on how to ameliorate HIV symptoms specifically for people living with HIV with HANA conditions; the VIP-HANA app therefore represents a first step in the development and dissemination of a consumer health informatics tool that incorporates evidence-based self-care strategies.

One of the major themes related to perceived ease of use in this study was avatar personalization. Within both group versions of the VIP-HANA app, participants were guided by an avatar through a series of questions ascertaining the nature and severity of their symptoms. The avatar then recommended self-care strategies for addressing the intervention group participants' symptoms. Even though the avatar in the control group version of the app did not provide any suggestions after the symptom assessment questions, participants in both groups felt that they connected to the app simply by selecting an avatar that most looked like them. Current research suggests that HIV patients feel more comfortable receiving HIV-related questions/answers from avatars compared to receiving health information directly from clinical providers, and users tend to choose an avatar that matches their own ethnicity [28,29]. Given that avatars can encourage user interaction with health-related information, developers of consumer health informatics tools should consider the potential benefits of inclusion of customizable avatars.

Privacy/confidentiality was a factor related to perceived ease of use. Our participants noted that the information (eg, data of participants' experienced symptoms) displayed on an app was secure because the app was installed on their own smartphone, and the app's design did not disclose their HIV status; thus, they felt comfortable using the app everywhere. In designing the app, privacy and confidentiality were strong considerations since people living with HIV in our formative work emphasized their importance [30] and social stigma associated with HIV persists [31]. Findings from our study emphasize the value of using mobile-based interventions, particularly for people living with HIV, and the importance of a user-centered design process with inclusion of users in terms of ensuring data security/protection.

Control group participants highlighted the need for inclusion of symptom reports in order to share their information with clinical providers without needing to memorize their symptom change/progress during the past 6 months. This functionality was only available for intervention group participants. Visual communication displayed on a summary on the app may improve patient comprehension of health information [32,33]. Findings from this work showed that graphical symptom reports enhanced the functionality of the app, which is an important implication for the future development of mHealth apps.

Limitations

There are some limitations to the generalizability of our results. Study participants who participated in the focus groups might have been those who were most likely to be engaged with and used the app who had been retained in the 6-month trial. More than half of our study participants were men, African American/Black, those who had completed some high school or less, and those who reported an annual income of less than

US $20,000. Therefore, the Hawthorne effect might influence the study findings of focus groups [34]. We designed the semistructured focus group guide based on our app's features and functionality, and used open-ended questions to minimize/eliminate concern about the Hawthorne effect.

Conclusions

Understanding factors influencing the use of patient-centered tools is critical. This study highlights participants' experiences, attitudes, and perceptions of their use of a behavioral intervention evaluated in a clinical trial. Participants were able to integrate the app into their daily routine, and used the app to support, track, monitor, and self-manage HIV symptoms related to HANA conditions. Although perceived ease of use was similar in both study groups, perceived usefulness differed between the groups. Participants in both groups found the VIP-HANA app to be useful in monitoring their symptoms of HANA conditions and enhancing communication with their clinical providers over time, but only the intervention group participants perceived the app to be useful for improving overall health and long-term symptom management in the HIV care continuum. Findings from this study highlight factors that are essential to ensure the usefulness of self-management apps to promote sustained use of the apps for people living with chronic illnesses.

Acknowledgments

This research was supported by the National Institute of Nursing Research, National Institutes of Health under award number R01 NR015737 (principal investigator RS) and through award K24NR018621 (principal investigator RS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Nursing Research, National Institutes of Health. We are grateful to Dr. Suzanne Bakken for leading the focus group sessions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of the VIP-HANA app features in the intervention and control group versions.

[DOCX File , 459 KB - jmir_v22i6e17708_app1.docx]

References

1. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
2. Meetoo D, Rylance R, Abuhaimid HA. Health care in a technological world. Br J Nurs 2018 Nov 08;27(20):1172-1177. [doi: 10.12968/bjon.2018.27.20.1172] [Medline: 30418859]
3. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User Preferences and Design Recommendations for an mHealth App to Promote Cystic Fibrosis Self-Management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
4. Wilhide Iii CC, Peeples MM, Anthony Kouyaté RC. Evidence-Based mHealth Chronic Disease Mobile App Intervention Design: Development of a Framework. JMIR Res Protoc 2016 Feb 16;5(1):e25 [FREE Full text] [doi: 10.2196/resprot.4838] [Medline: 26883135]
5. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
6. Spirig R, Moody K, Battegay M, De Geest S. Symptom management in HIV/AIDS: advancing the conceptualization. ANS Adv Nurs Sci 2005;28(4):333-344. [doi: 10.1097/00012272-200510000-00005] [Medline: 16292019]
7. Schnall R, Bakken S, Rojas M, Travers J, Carballo-Dieguez A. mHealth Technology as a Persuasive Tool for Treatment, Care and Management of Persons Living with HIV. AIDS Behav 2015 Jun 9;19(S2):81-89 [FREE Full text] [doi: 10.1007/s10461-014-0984-8] [Medline: 25572830]
8. Stonbraker S, Cho H, Hermosi G, Pichon A, Schnall R. Usability Testing of a mHealth App to Support Self-Management of HIV-Associated Non-AIDS Related Symptoms. Stud Health Technol Inform 2018;250:106-110 [FREE Full text] [Medline: 29857399]
9. Schnall R, Liu J, Mohr DC, Bakken S, Hirshfield S, Siegel K, et al. Multi-Modal Methodology for Adapting Digital Health Tools to New Populations: Adaptation of the Video Information Provider (VIP) for Persons Living with HIV with HIV-Associated Non-AIDS (HANA) Conditions. Stud Health Technol Inform 2019 Aug 21;264:1347-1351. [doi: 10.3233/SHTI190446] [Medline: 31438145]
10. Iribarren S, Siegel K, Hirshfield S, Olender S, Voss J, Krongold J, et al. Self-Management Strategies for Coping with Adverse Symptoms in Persons Living with HIV with HIV Associated Non-AIDS Conditions. AIDS Behav 2018 Jan 9;22(1):297-307 [FREE Full text] [doi: 10.1007/s10461-017-1786-6] [Medline: 28488165]

11. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
12. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. BMJ 2009 Sep 10;339:b3496 [FREE Full text] [doi: 10.1136/bmj.b3496] [Medline: 19744976]
13. Rapport F, Storey M, Porter A, Snooks H, Jones K, Peconi J, et al. Qualitative research within trials: developing a standard operating procedure for a clinical trials unit. Trials 2013 Feb 21;14:54 [FREE Full text] [doi: 10.1186/1745-6215-14-54] [Medline: 23433341]
14. Henry G. Applied Social Research Methods: Practical sampling. Thousand Oaks, CA: SAGE Publications, Inc; 1990.
15. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quart 1989 Sep;13(3):319. [doi: 10.2307/249008]
16. Swain J. A Hybrid Approach to Thematic Analysis in Qualitative Research: Using a Practical Example. London: SAGE Research Methods Cases; 2018.
17. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS Med 2005 Oct;2(10):e267 [FREE Full text] [doi: 10.1371/journal.pmed.0020267] [Medline: 16138788]
18. David FN, Tukey JW. Exploratory Data Analysis. Boston: Addison-Wesley Pub Co; Dec 1977.
19. Simoni JM, Pearson CR, Pantalone DW, Marks G, Crepaz N. Efficacy of interventions in improving highly active antiretroviral therapy adherence and HIV-1 RNA viral load. A meta-analytic review of randomized controlled trials. J Acquir Immune Defic Syndr 2006 Dec 01;43(Suppl 1):S23-S35 [FREE Full text] [doi: 10.1097/01.qai.0000248342.05438.52] [Medline: 17133201]
20. Tzaneva V, Iacob T. A better communcation with the patients improves the management of HIV disease: a nonsystematic review. Clujul Med 2013;86(3):181-184 [FREE Full text] [Medline: 26527943]
21. Decroo T, Rasschaert F, Telfer B, Remartinez D, Laga M, Ford N. Community-based antiretroviral therapy programs can overcome barriers to retention of patients and decongest health services in sub-Saharan Africa: a systematic review. Int Health 2013 Sep 30;5(3):169-179. [doi: 10.1093/inthealth/iht016] [Medline: 24030268]
22. Sharma R, Mishra R. A Review of Evolution of Theories and Models of Technology Adoption. Indore Management J 2014;6(2):17-29.
23. Isaac O, Abdullah Z, Ramayah T, Mutahar A, Alrajawy I. Perceived Usefulness, Perceived Ease of Use, Perceived Compatibility, and Net Benefits: an empirical study of internet usage among employees in Yemen. 2016 Presented at: 7th International Conference Postgraduate Education (ICPE7); December 1, 2016; Universiti Teknologi MARA (UiTM), Malaysia p. 899-919.
24. Reppa I, McDougall S. When the going gets tough the beautiful get going: aesthetic appeal facilitates task performance. Psychon Bull Rev 2015 Oct 17;22(5):1243-1254 [FREE Full text] [doi: 10.3758/s13423-014-0794-z] [Medline: 25595208]
25. Wantland DJ, Holzemer WL, Moezzi S, Willard SS, Arudo J, Kirksey KM, et al. A randomized controlled trial testing the efficacy of an HIV/AIDS symptom management manual. J Pain Symptom Manage 2008 Sep;36(3):235-246 [FREE Full text] [doi: 10.1016/j.jpainsymman.2007.10.011] [Medline: 18400461]
26. Cho H, Porras T, Baik D, Beauchemin M, Schnall R. Understanding the predisposing, enabling, and reinforcing factors influencing the use of a mobile-based HIV management app: A real-world usability evaluation. Int J Med Inform 2018 Sep;117:88-95 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.007] [Medline: 30032969]
27. Iribarren S, Siegel K, Hirshfield S, Olender S, Voss J, Krongold J, et al. Self-Management Strategies for Coping with Adverse Symptoms in Persons Living with HIV with HIV Associated Non-AIDS Conditions. AIDS Behav 2018 Jan 9;22(1):297-307 [FREE Full text] [doi: 10.1007/s10461-017-1786-6] [Medline: 28488165]
28. Canidate S, Hart M. The Use of Avatar Counseling for HIV/AIDS Health Education: The Examination of Self-Identity in Avatar Preferences. J Med Internet Res 2017 Dec 01;19(12):e365 [FREE Full text] [doi: 10.2196/jmir.6740] [Medline: 29196281]
29. Gordon R, Björklund NK, Smith RJ, Blyden ER. Halting HIV/AIDS with avatars and havatars: a virtual world approach to modelling epidemics. BMC Public Health 2009 Nov 18;9(S1):S13. [doi: 10.1186/1471-2458-9-s1-s13]
30. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, Perceived Risk, Perceived Ease of Use and Perceived Usefulness as Factors Related to mHealth Technology Use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]
31. Anderson B. HIV Stigma and Discrimination Persist, Even in Health Care. Virtual Mentor 2009 Dec 01;11(12):998-1001 [FREE Full text] [doi: 10.1001/virtualmentor.2009.11.12.oped1-0912] [Medline: 23207098]
32. Siricharoen WV. Infographic role in helping communication for promoting health and well-being. 2015 Presented at: Second international conference on computer science, computer engineering, and education technologies (CSCEET2015); September 8-10, 2015; Kuala Lumpur, Malaysia.
33. Dowse R, Ramela T, Browne SH. An illustrated leaflet containing antiretroviral information targeted for low-literate readers: development and evaluation. Patient Educ Couns 2011 Dec;85(3):508-515. [doi: 10.1016/j.pec.2011.01.013] [Medline: 21306856]

34. Holden JD. Hawthorne effects and research into professional practice. J Eval Clin Pract 2001 Feb 07;7(1):65-70. [doi: 10.1046/j.1365-2753.2001.00280.x] [Medline: 11240840]

Abbreviations

HANA: HIV-associated nonAIDS
mHealth: mobile health
RCT: randomized controlled trial
TAM: technology acceptance model
VIP-HANA: video information provider for HIV-associated nonAIDS conditions

Edited by G Eysenbach; submitted 07.01.20; peer-reviewed by YC Kato-Lin, V Strotbaum; comments to author 10.03.20; revised version received 07.04.20; accepted 21.04.20; published 15.06.20.

Please cite as:
Cho H, Porras T, Flynn G, Schnall R
Usability of a Consumer Health Informatics Tool Following Completion of a Clinical Trial: Focus Group Study
J Med Internet Res 2020;22(6):e17708
URL: http://www.jmir.org/2020/6/e17708/
doi:10.2196/17708
PMID:32538796

Original Paper

Facebook Recruitment Using Zip Codes to Improve Diversity in Health Research: Longitudinal Observational Study

Cornelia Pechmann[1], BSc, MBA, PhD; Connor Phillips[1], BA; Douglas Calder[1], BA; Judith J Prochaska[2], MPH, PhD

[1]Paul Merage School of Business, University of California, Irvine, Irvine, CA, United States

[2]Stanford Prevention Research Center, Department of Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Cornelia Pechmann, BSc, MBA, PhD
Paul Merage School of Business
University of California, Irvine
School of Business I
Irvine, CA,
United States
Phone: 1 949 824 4058
Email: cpechman@uci.edu

Abstract

Background: Facebook's advertising platform reaches most US households and has been used for health-related research recruitment. The platform allows for advertising segmentation by age, gender, and location; however, it does not explicitly allow for targeting by race or ethnicity to facilitate a diverse participant pool.

Objective: This study looked at the efficacy of zip code targeting in Facebook advertising to reach blacks/African Americans and Hispanics/Latinos who smoke daily for a quit-smoking web-based social media study.

Methods: We ran a general market campaign for 61 weeks using all continental US zip codes as a baseline. Concurrently, we ran 2 campaigns to reach black/African American and Hispanic-/Latino-identified adults, targeting zip codes ranked first by the percentage of households of the racial or ethnic group of interest and then by cigarette expenditure per household. We also ran a Spanish language campaign for 13 weeks, targeting all continental US zip codes but utilizing Facebook's Spanish language targeting. The advertising images and language were common across campaigns. Costs were compared for advertisement clicks, queries, applications, and participants, and yields were compared for the final three outcomes. We examined outcomes before and after the Cambridge Analytica scandal that broke in March 2018. Finally, we examined 2 promoted Facebook features: lookalike audiences and audience network placement.

Results: Zip code targeting campaigns were effective for yielding the racial or ethnic groups of interest. The black-/African American–focused versus general market campaign increased black/African American weekly queries (mean 9.48, SD 5.69 vs general market mean 2.83, SD 2.05; *P*<.001) and applicants (mean 1.11, SD 1.21 vs general market mean 0.54, SD 0.58; *P*<.001). The Hispanic-/Latino-focused versus general market campaign increased Hispanic/Latino weekly queries (mean 3.10, SD 2.16 vs general market mean 0.71, SD 0.48; *P*<.001) and applicants (mean 0.36, SD 0.55 vs general market mean 0.10, SD 0.14; *P*=.001). Cost metrics did not differ between campaigns at generating participants (overall *P*=.54). Costs increased post- versus prescandal for the black-/African American–focused campaign for queries (mean US $8.51, SD 3.08 vs US $5.87, SD 1.89; *P*=.001) and applicants (mean US $59.64, SD 35.63 vs US $38.96, SD 28.31; *P*=.004) and for the Hispanic-/Latino-focused campaign for queries (mean US $9.24, SD 4.74 vs US $7.04, SD 3.39; *P*=.005) and applicants (mean US $61.19, SD 40.08 vs US $38.19, SD 21.20; *P*=.001).

Conclusions: Zip code targeting in Facebook advertising is an effective way to recruit diverse populations for health-based interventions. Audience network placement should be avoided. The Facebook lookalike audience may not be necessary for recruitment, with drawbacks including an unknown algorithm and unclear use of Facebook user data, and so public concerns around data privacy should be considered.

Trial Registration: ClinicalTrial.gov NCT02823028; https://clinicaltrials.gov/ct2/show/NCT02823028

(J Med Internet Res 2020;22(6):e17554) doi:10.2196/17554

KEYWORDS

smoking; advertisement; social media

Introduction

Background

The distrust of health care and health-related studies among blacks/African Americans and Hispanics/Latinos has been documented extensively [1-3]. To combat racial and ethnic homogeneity in health-related research, emphasis must be placed on recruitment of diverse participants, especially in studies concerning diseases or products in which race or ethnicity is a factor [4,5]. Tobacco companies have traditionally advertised products to communities of color [6-8]. Their targeting methods have included placing advertisements on television shows, in print media, and on websites with high black/African American or Hispanic/Latino viewership [9,10]. Tobacco companies have also spatially targeted communities of color by placing billboards and bus advertisements in neighborhoods primarily comprising black/African American or Hispanic/Latino residents [11]. Tobacco companies have paid people to go into inner-city neighborhoods to hand out free samples of menthol (and sometimes regular) cigarettes in an effort to attract black/African American young adult and adult customers [12-14].

Facebook is a leading web-based social media platform for adults in the United States, with approximately 221 million monthly active users who represent about 69% of the adult population [15,16]. Facebook is also used at similar rates among whites, blacks/African Americans, and Hispanics/Latinos [17]. Facebook advertising has increasingly been used for health-related study recruitment [18]. Cost efficiency and widespread adult use, as well as detailed targeting features, make Facebook advertising a popular choice for study recruitment [18,19]. Participants recruited through the Facebook advertising platform were found to be demographically similar compared with traditional study recruitment methods, such as print materials [18,20]. In some cases, however, participant pools showed an overrepresentation of non-Hispanic white individuals [21]. Race and ethnicity are important aspects of study recruitment, and more emphasis is being placed on inclusion and reporting of racial or ethnic diversity in the participant pool, especially in health-related research [22]. However, there remains little discussion about how to target diverse racial or ethnic groups using Facebook advertising [23].

This paper discusses recruitment of racially and ethnically diverse participants by means of Facebook advertising for a quit-smoking web-based social media study. Our research team developed a web-based Twitter peer-support group intervention for quitting smoking called Tweet2Quit [24,25]. We recruited virtually all participants (N=980) using Facebook advertising and sought an ethnically diverse participant pool.

Study Goals

This study evaluated 4 Facebook advertising campaigns that we used to recruit individuals for our quit-smoking web-based social media study. Our general market campaign reached all continental US zip codes, targeting individuals who expressed an interest in smoking or quitting smoking on Facebook, and it served as our baseline. Two additional Facebook campaigns targeted primarily black/African American or Hispanic/Latino zip codes with high cigarette expenditures. Our final Facebook campaign targeted Spanish language speakers living in continental US zip codes who expressed an interest in smoking or quitting smoking on Facebook. For each campaign, we examined yields and costs for 4 standard advertising campaign outcomes: advertisement clicks, queries, applications, and participants. We hypothesized that zip code campaigns would generate higher yields of diverse ethnic groups without higher costs.

In addition, we compared costs before and after major news broke in March 2018 concerning the Facebook privacy scandal involving Cambridge Analytica’s alleged breach of privacy of 50 million Facebook users, which raised serious concerns about the protection of private user information. Finally, we examined the costs associated with 2 features Facebook recommends to its advertisers, use of lookalike audiences and audience network placement, and compared these to our general market campaign.

Methods

Facebook Advertising Campaigns

Facebook does not allow for direct demographic targeting of advertising by users’ ethnicity or race. The main demographics that Facebook advertisers can choose are age and gender. These demographics are offered because when Facebook users create their accounts, they are required to specify their age and gender. Advertisers can further narrow their audience by using the 3 sections labeled *Demographics*, *Interests*, and *Behaviors*. The *Demographics* and *Interests* sections do not include race or ethnicity. The *Behaviors* section includes Multicultural Affinity, which allows advertisers to choose African American, Asian American, or Hispanic affinity. Facebook does not provide information on how individuals are classified into these behavioral affinity categories or what data are used. In addition, categories within the *Interests* and *Behaviors* sections change often, which then requires advertisers to recreate existing advertising campaigns using new options. Furthermore, advertising using Multicultural Affinity is restricted by Facebook and can lead to extended advertising review times and automated advertisement disapproval [26]. Due to these reasons, we used more standard zip code targeting in our study.

As a baseline, we created a general market campaign, selecting all continental US zip codes. The advertising images and language were common in all 4 advertising campaigns we studied. In addition, in all 4 campaigns, we restricted our advertisements to be seen by individuals aged between 21-59 years. We required individuals to be ≥21 years because many states restrict the sale of tobacco to these individuals. We set the maximum age at 59 to focus on younger and middle-aged adults rather than retirees. Moreover, in all 4 of our advertising campaigns, we targeted appropriate study participants based on their Facebook interests. We showed our advertisements only to individuals who expressed an interest in smoking-related topics such as quitting smoking, nicotine, and tobacco or cigarettes, while logged into Facebook.

For our black-/African American–focused and Hispanic-/Latino-focused campaigns, we also used more specific zip code targeting. We obtained data on all continental US zip codes and sorted the zip codes from highest to lowest based on the percentage of households in the target racial or ethnic group to ensure that those racial or ethnic groups were reached. After this, we sorted the zip codes from highest to lowest based on the mean annual household expenditure on cigarettes to try to ensure that we reached smokers. We removed any zip codes with fewer than 100 households. We then selected the top 1000 zip codes because Facebook only allowed us to upload 1000 customized zip codes out of the 41,702 total zip codes in the United States [27-29]. All chosen zip codes had high percentages of ethnic households and high cigarette expenditures. The 1000 zip codes we used for the black-/African American–focused campaign ranged from 53% to 99% in terms of the households in this ethnic group (mean 70%, SD 0.13%), with annual cigarette expenditures from US $154 to US $568 (mean US $329, SD 62.49) per household. The 1000 zip codes we used for the Hispanic-/Latino-focused campaign ranged from 53% to 99% (mean 71%, SD 0.13%) in terms of the households in this ethnic group, with annual cigarette expenditures from US $70 to US $506 (mean US $266, SD 69.86) per household. We uploaded the targeted zip codes for each campaign on Facebook.

Finally, for our Spanish language campaign, we used all continental US zip codes; however, we selected Spanish as the spoken language, using Facebook's designated Language Targeting feature. This feature allows advertisers to show advertisements to users who speak a specific language and is not related to targeting based on the *Behaviors* or *Interests* sections. Spanish language targeting was chosen to increase the number of Hispanic/Latino applicants to our program. Advertisements across all campaigns were identical, as noted above. Advertising imagery and wording were chosen using Facebook A/B testing campaigns placed in all US zip codes. A/B testing allows advertisers to run nearly identical, simultaneous campaigns to test a singular variable, such as an advertising image.

Data on the ethnic makeup of US zip codes came from the US Census Bureau's American Community Survey and are publicly available [27]. Data on cigarette expenditures came from the US Bureau of Labor Statistics' Consumer Expenditure Survey and, although not publicly available at the zip code level, can be purchased from Experian [28,30].

Outcome Measures

Our study compared 4 different Facebook advertising campaigns: A general market campaign, a black-/African American–focused zip code campaign, a Hispanic-/Latino-focused zip code campaign, and a Spanish language campaign. We compared these campaigns on 4 standard outcome measures: advertisement clicks, queries, applicants, and participants. Our advertisement click measure showed whether an individual tapped or clicked on our web-based advertisement, which automatically directed the individual to our study website. Hence, our advertisement click measure assessed if a campaign possibly brought interested individuals to our website. On our website, individuals could read about the study and fill out a short query form that asked for their contact information (name and email) and their race or ethnicity. Our query measure assessed whether the campaign resulted in individuals providing contact information. Individuals who provided contact information received our full screening survey, and if they finished the survey, they were considered applicants. Hence, our applicant measure assessed whether the campaign resulted in individuals completing the study screener, regardless of whether they passed or failed. Finally, our participant measure assessed whether the campaign resulted in individuals getting enrolled into the study.

Our initial Facebook advertising testing began in October 2016 and included testing of the Facebook-recommended audience network placement. We began recruitment for our study in January 2017 but did not commence our sophisticated Pixel-based measurement of our Facebook campaigns (explained below) until mid-June 2017 and that continued until September 2018 for a total of 61 weeks. The Facebook scandal occurred in March 2018 (Multimedia Appendix 1).

Descriptive Statistics on Overall Campaign Response

Across our 4 Facebook campaigns, we received a total of 92,677 advertisement clicks between June 20, 2017, and September 9, 2018. The general market campaign received 66,681 advertisement clicks, the black-/African American–focused campaign received 12,544 advertisement clicks, and the Hispanic-/Latino-focused campaign received 10,969 advertisement clicks. Our shorter Spanish language campaign received 2483 advertisement clicks between June 20, 2017, and September 30, 2017. On average, about 13.92% (12,898/92,677) of individuals who clicked our advertisement filled out the interest form, constituting a query. After receiving an email with a link to our web-based screening survey, about 19.31% (2490/12,898) of individuals completed the survey and about 20.32% (506/2490) of these applicants were enrolled in our study. We could not capture ethnicity at the advertisement click stage because we had no mechanism for doing this. However, we began collecting self-reported race or ethnicity at the query stage and continued this through the applicant and participant stages.

We examined whether study exclusion was related to the individual's race or ethnicity. Race or ethnicity related to study exclusion due to a health contraindication (eg, pregnancy), in that blacks/African Americans (69/259, 26.6%) were more likely to be excluded for health reasons than non-Hispanic whites (295/1544, 19.11%; X^2_1=7.8; P=.005). Ethnicity related to study exclusion because of refusing mobile phone verification, in that whites (308/1544, 19.95%) were more likely to be excluded for this reason than blacks/African Americans (36/259, 13.9%; X^2_1=5.3; P=.02). Ethnicity related to study exclusion because of smoking too few cigarettes per day to permit use of the study-provided nicotine replacement therapy, in that blacks/African Americans (11/259, 4.2%) were more likely to be excluded for this reason than whites (28/1544, 1.81%;

X^2_1=6.2; P=.01). No other effects were significant (all P>.17; Multimedia Appendix 2).

Results

Results Regarding Yields by Racial or Ethnic Group

Overview of Yield Analysis

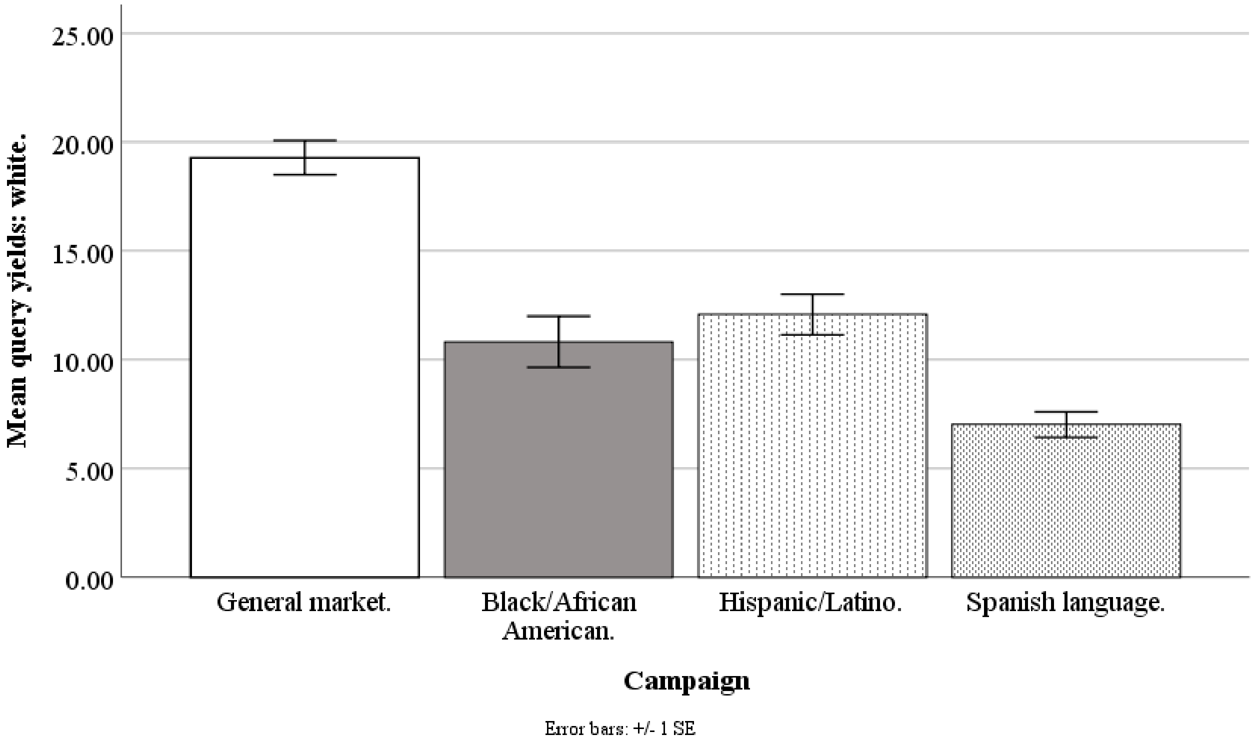
Facebook's basic advertisement reporting system measures advertisement clicks by default. To obtain the query, applicant, and participant metrics, we utilized a Facebook Pixel, which is a web code our website developer installed to measure and track actions Facebook users performed on our study website. The Facebook Pixel also tracked which of our Facebook campaigns drew them to our website [31]. We installed the Facebook Pixel in mid-June 2017. Our yield analysis is based on weekly data from June 20, 2017, to September 30, 2018, excluding major holidays (n=61 weeks), except Spanish language data that are from June 20, 2017, to September 30, 2017 (n=15 weeks). Although the Facebook Ads Manager provides daily yields, we aggregated the daily data to weekly data for accuracy and smoothing (eg, staff did not work weekends to convert queries to applicants). For all statistical analyses, we used analysis of variance, after which we conducted two-tailed pairwise *t* tests that compared specific campaigns (eg, general market vs Spanish language) using the Sidak correction for multiple comparisons.

We could not directly compare campaign yields because expenditures differed by campaign (eg, our budget for the general market campaign was considerably higher than for the black-/African American–focused and Hispanic-/Latino-focused campaigns). To compare yields across campaigns, we standardized expenditures to US $140 per week (US $20 per day), reflecting our average budget for the 2 main ethnic campaigns. For example, on a given week, we may have spent US $700 on the general market campaign versus US $140 on the black-/African American–focused campaign. If we received 100 queries from the general market campaign, we divided this number by 5 and estimated the yield as 20 based on a standardized expenditure of US $140 per week (US $700/5=US $140). This standardized yield of 20 for the general market campaign would be compared with the yield of the black-/African American–focused campaign that ran at US $140.

Non-Hispanic White Query Yields

The Facebook campaigns differed on weekly mean counts of queries by non-Hispanic whites ($F_{3,194}$=19.79; P<.001), with the general market–focused campaign at 19.28 people (SD 6.11), black-/African American–focused at 10.82 people (SD 9.20), Hispanic-/Latino-focused at 12.10 people (SD 7.25), and Spanish language–focused at 7.02 people (SD 2.24). The query yield of whites was higher for the general market campaign than for the black-/African American–focused ($t_{1,194}$=6.33; P<.001), Hispanic-/Latino-focused ($t_{1,194}$=5.39; P<.001), or Spanish language–focused campaigns ($t_{1,194}$=5.77; P<.001). The query yield of whites was similar for the black-/African American–focused campaign versus Hispanic-/Latino-focused ($t_{1,194}$=0.94; P=.92) or Spanish language campaigns ($t_{1,194}$=1.79; P=.37). The weekly mean count of whites was similar for the Hispanic-/Latino-focused versus Spanish language campaign ($t_{1,194}$=2.38; P=.11; Figure 1).

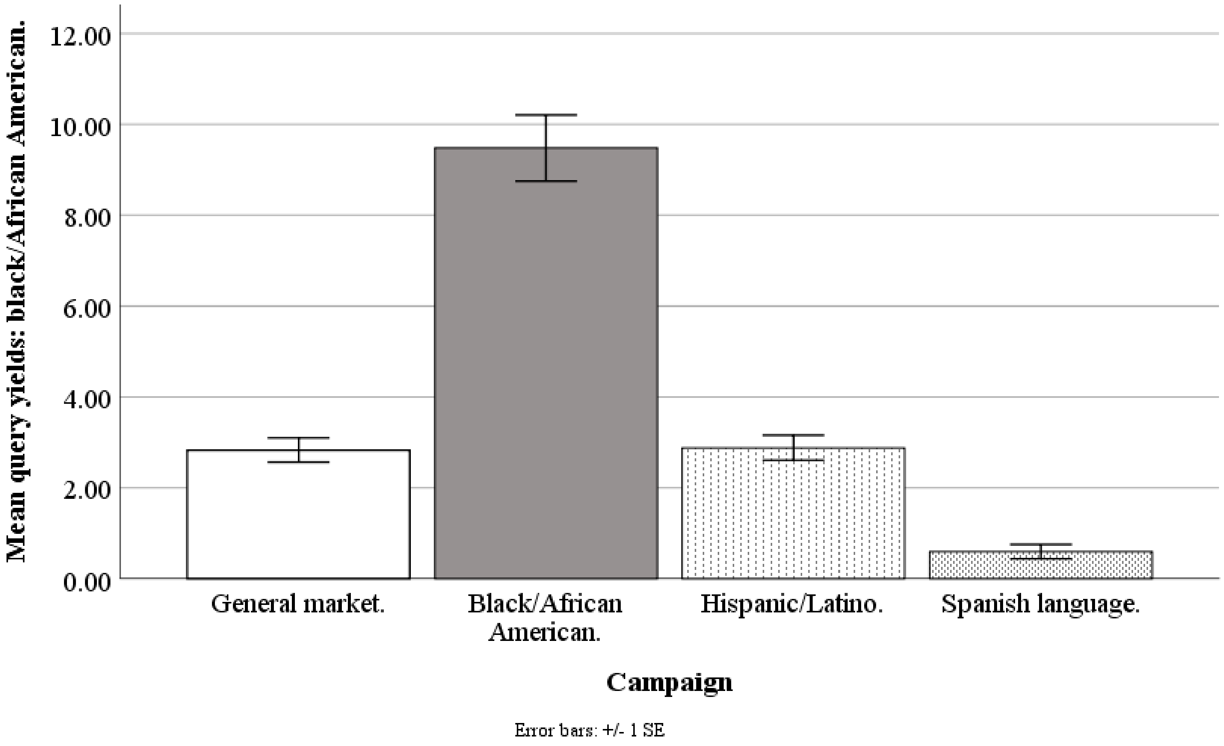
Figure 1. Mean query yields of whites for Facebook advertising campaigns.



Black/African American Query Yield

The Facebook campaigns differed on weekly mean counts of queries by blacks/African Americans ($F_{3,194}$=53.96; P<.001), with the general market–focused campaign at 2.83 people (SD 2.05), black-/African American–focused at 9.48 people (SD 5.69), Hispanic-/Latino-focused at 2.89 people (SD 2.12), and Spanish language–focused at 0.60 people (SD 0.62). The query yield of blacks/African Americans was higher for the black-/African American–focused campaign than for the general market–focused ($t_{1,194}$=10.29; P<.001), Hispanic-/Latino-focused ($t_{1,194}$=10.21; P<.001), or Spanish language–focused campaigns ($t_{1,194}$=8.64; P<.001). The query yield of blacks/African Americans was similar for the general market–focused versus Hispanic-/Latino-focused ($t_{1,194}$=0.08; P=.99) or Spanish language–focused campaigns ($t_{1,194}$=2.17; P=.17). The query yield of blacks/African Americans was similar for the Hispanic-/Latino-focused versus Spanish language–focused campaign ($t_{1,194}$=2.23; P=.15; Figure 2).

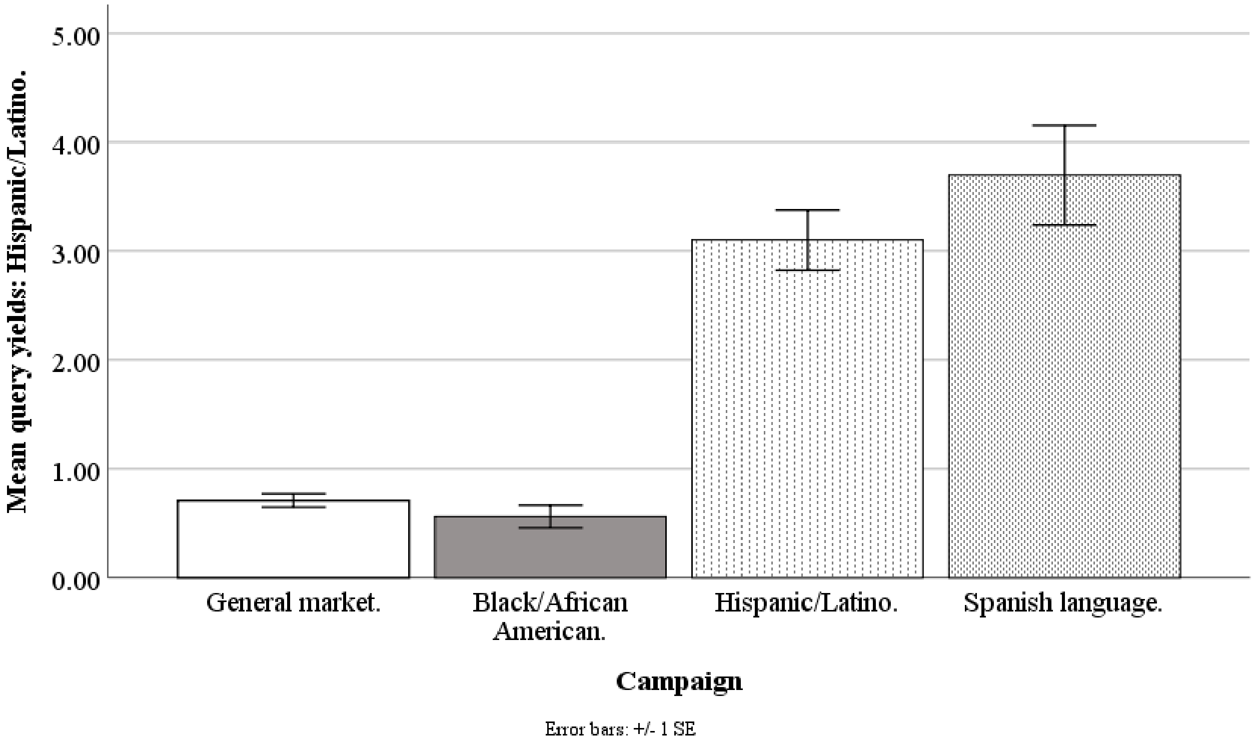
Figure 2. Mean query yields of blacks/African Americans for Facebook advertising campaigns.



Hispanic/Latino Query Yield

The Facebook campaigns differed on weekly mean counts of queries by Hispanics/Latinos ($F_{3,194}$=54.29; P<.001), with the general market–focused campaign at 0.71 people (SD 0.48), black-/African American–focused at 0.56 people (SD 0.82), Hispanic-/Latino-focused at 3.10 people (SD 2.16), and Spanish language–focused at 3.70 people (SD 1.77). The query yield of Hispanics/Latinos was higher for the Hispanic-/Latino-focused campaign than for the general market–focused ($t_{1,194}$=9.45; P<.001) or black-/African American–focused campaigns ($t_{1,194}$=10.04; P<.001) but was similar to the Spanish language–focused campaign ($t_{1,194}$=1.48; P=.60). The query yield of Hispanics/Latinos was similar for the general market–focused campaign as compared with the black-/African American–focused campaign ($t_{1,194}$=0.58; P=.99) but was lower than that of the Spanish language–focused campaign ($t_{1,194}$=7.43; P<.001). The query yield of Hispanics/Latinos was higher for the Spanish language–focused versus African American–focused campaign ($t_{1,194}$=7.79; P<.001; Figure 3).

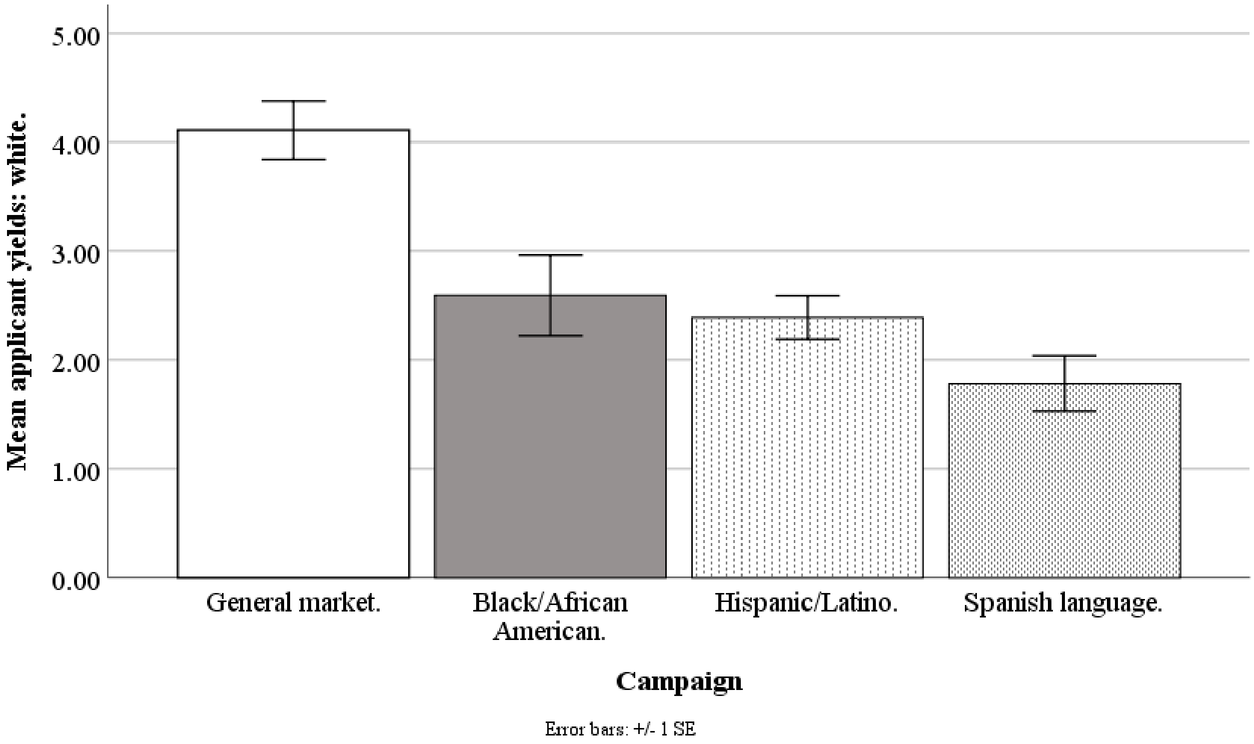
Figure 3. Mean query yields of Hispanics/Latinos for Facebook advertising campaigns.



Non-Hispanic White Applicant Yield

The Facebook campaigns differed on weekly mean counts of applicants who were non-Hispanic white ($F_{3,190}$=9.00; P<.001), with the general market–focused campaign at 4.12 (SD 2.08), black-/African American–focused at 2.59 (SD 2.86), Hispanic-/Latino-focused at 2.39 (SD 1.55), and Spanish language–focused at 1.78 (SD 0.95). The applicant yield of whites was higher for the general market–focused campaign than for the black-/African American–focused ($t_{1,190}$=3.84; P=.001), Hispanic-/Latino-focused ($t_{1,190}$=4.35; P<.001), or Spanish language–focused campaigns ($t_{1,190}$=3.62; P=.002). The applicant yield of whites was similar for the black-/African American–focused versus Hispanic-/Latino-focused ($t_{1,190}$=0.51; P=.99) or Spanish language–focused campaigns ($t_{1,190}$=1.26; P=.76). The applicant yield of whites was similar for the Hispanic-/Latino-focused versus Spanish language–focused campaign ($t_{1,190}$=0.95; P=.92; Figure 4).

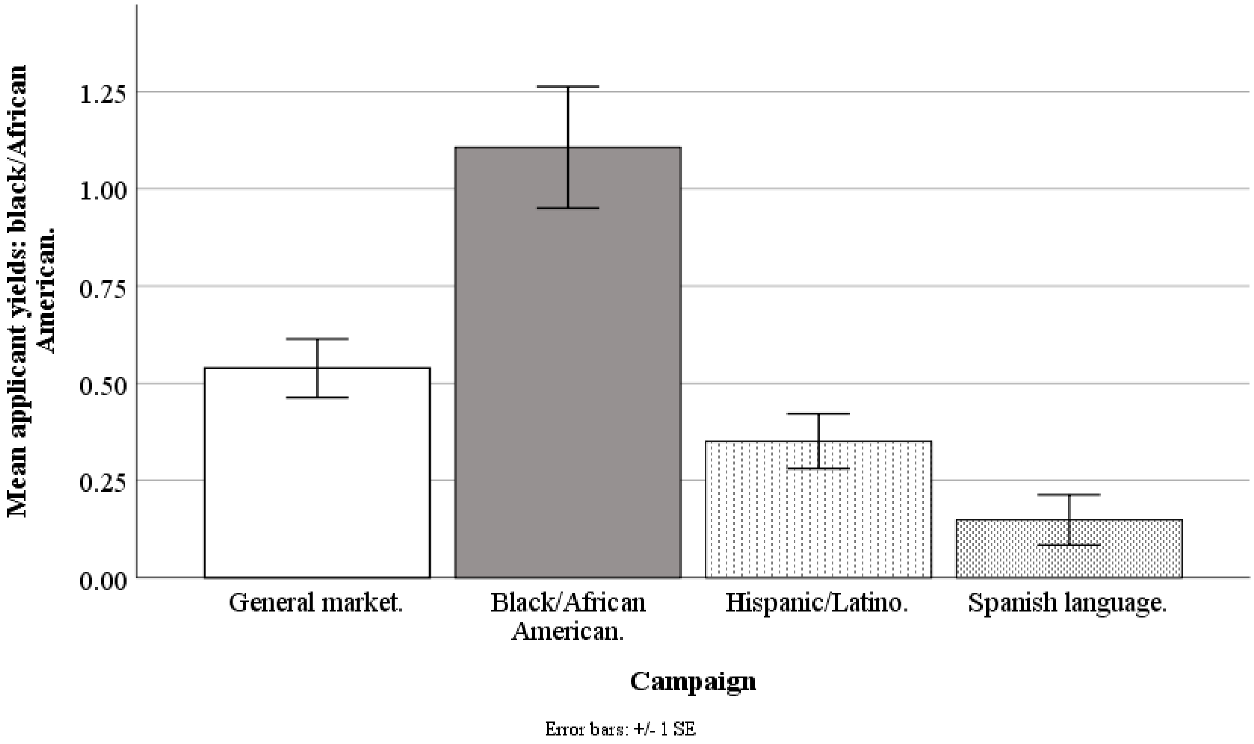
Figure 4. Mean applicant yields of whites for Facebook advertising campaigns.



Black/African American Applicant Yields

The Facebook campaigns differed on weekly mean counts of applicants who were black/African American ($F_{3,190}$=11.19; P<.001), with the general market–focused campaign at 0.54 (SD 0.58), black-/African American–focused at 1.11 (SD 1.21), Hispanic-/Latino-focused at 0.35 (SD 0.55), and Spanish language–focused at 0.15 (SD 0.24). The applicant yield of blacks/African Americans was higher for the black-/African American–focused campaign compared with the general market–focused ($t_{1,190}$=3.84; P=.001), Hispanic-/Latino-focused ($t_{1,190}$=5.11; P<.001), or Spanish language–focused campaigns ($t_{1,190}$=3.96; P=.001). The applicant yield of blacks/African Americans was similar for the general market–focused versus Hispanic-/Latino-focused ($t_{1,190}$=1.27; P=.75) or Spanish language–focused campaigns ($t_{1,190}$=1.62; P=.49). The applicant yield of blacks/African Americans was similar for the Hispanic-/Latino-focused versus Spanish language–focused campaign ($t_{1,190}$=0.84; P=.95; Figure 5).

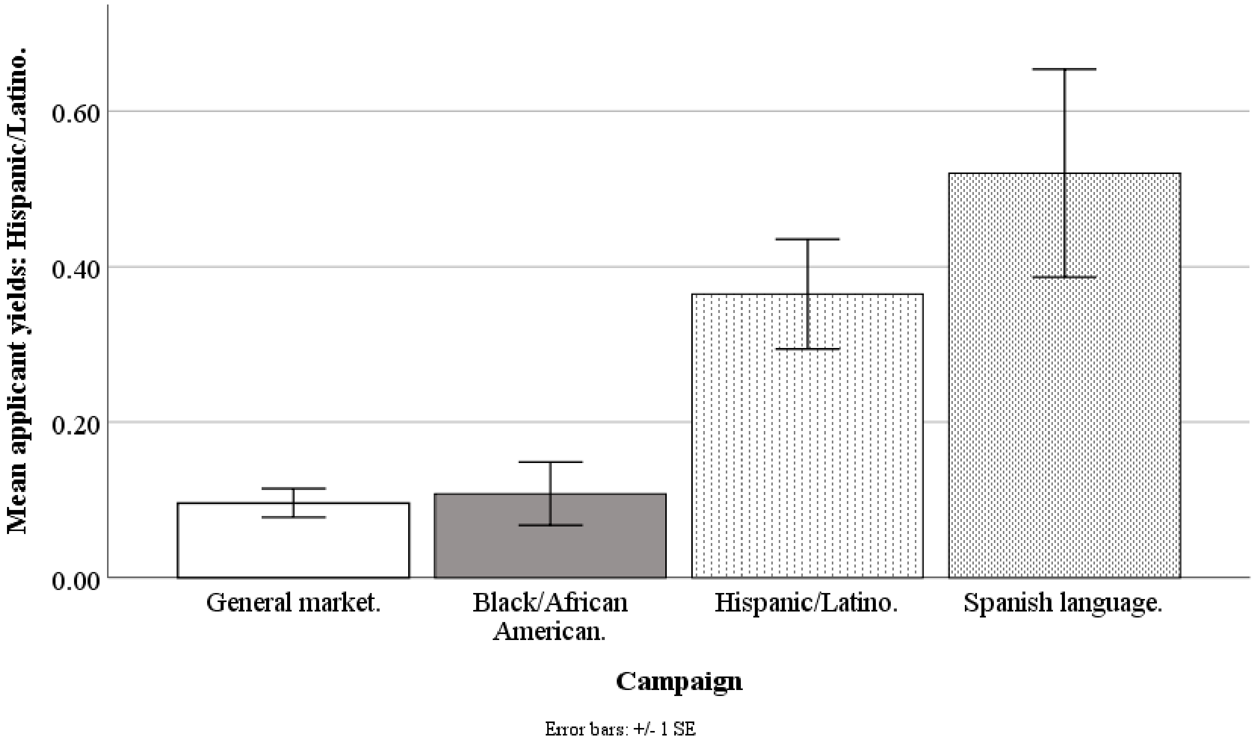
Figure 5. Mean applicant yields of blacks/African Americans for Facebook advertising campaigns.



Hispanic/Latino Applicant Yield

The Facebook campaigns differed on weekly mean counts of applicants who were Hispanic/Latino ($F_{3,190}$=9.48; *P*<.001), with the general market–focused campaign at 0.10 (SD 0.14), black-/African American–focused at 0.11 (SD 0.32), Hispanic-/Latino-focused at 0.36 (SD 0.55), and Spanish language–focused at 0.52 (SD 0.50). The applicant yield of Hispanics/Latinos was higher for the Hispanic-/Latino-focused campaign compared with general market–focused ($t_{1,190}$=3.83; *P*=.001) or black-/African American–focused campaigns ($t_{1,190}$=3.66; *P*=.002) but similar to that of Spanish language–focused campaign ($t_{1,190}$=1.37; *P*=.68). The applicant yield of Hispanics/Latinos was similar for the general market–focused campaign compared with the black-/African American–focused campaign ($t_{1,190}$=0.17; *P*=.99) but lower than that of the Spanish language–focused campaign ($t_{1,190}$=3.72; *P*=.002). The applicant yield of Hispanics/Latinos was higher for the Spanish language–focused campaign versus the black-/African American-focused campaign ($t_{1,190}$=3.61; *P*=.002; Figure 6).

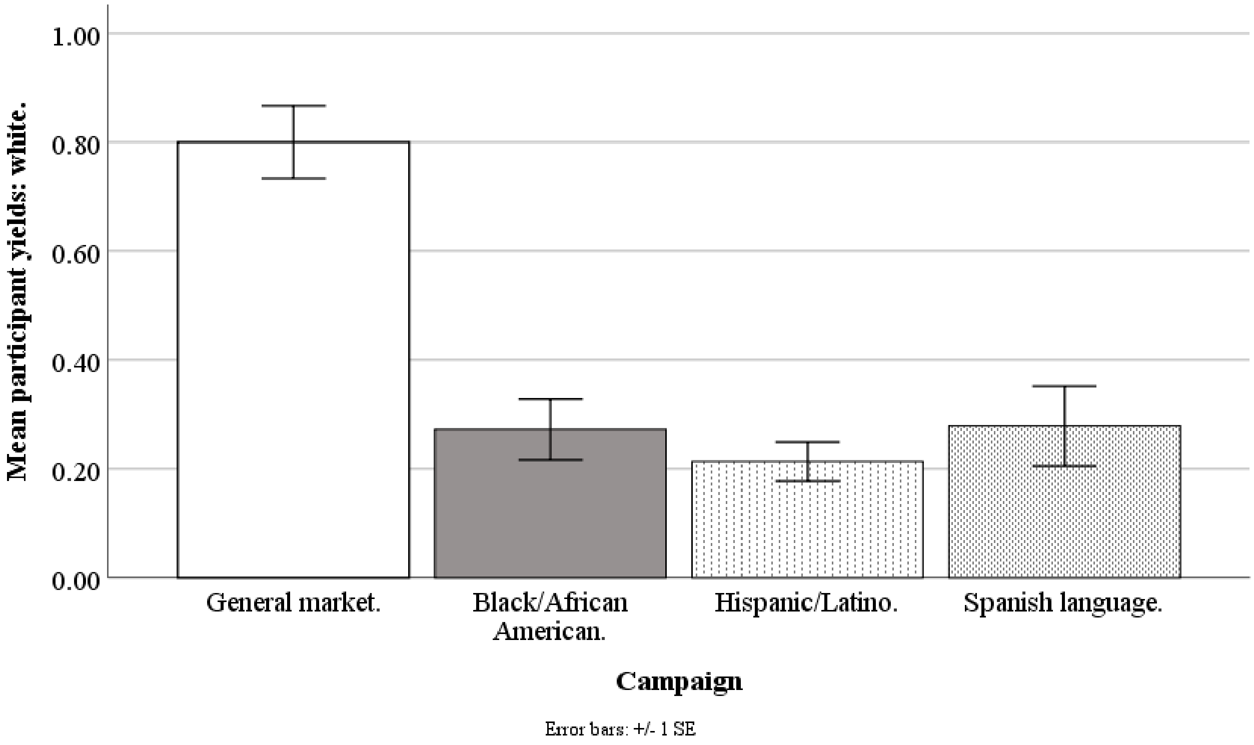
Figure 6. Mean applicant yields of Hispanics/Latinos for Facebook advertising campaigns.



Non-Hispanic White Participant Yields

The Facebook campaigns differed on weekly mean counts of participants who were non-Hispanic white ($F_{3,190}$=24.96; $P<.001$), with the general market–focused campaign at 0.80 (SD 0.52), black-/African American–focused at 0.27 (SD 0.43), Hispanic-/Latino-focused at 0.21 (SD 0.28), and Spanish language–focused at 0.28 (SD 0.28). The participant yield of whites was higher for the general market–focused campaign than for the black-/African American–focused ($t_{1,190}$=6.95; $P<.001$), Hispanic-/Latino-focused ($t_{1,190}$=7.72; $P<.001$), or Spanish language–focused campaigns ($t_{1,190}$=4.24; $P<.001$). The participant yield of whites otherwise did not differ across campaigns (all $P>.97$; Figure 7).

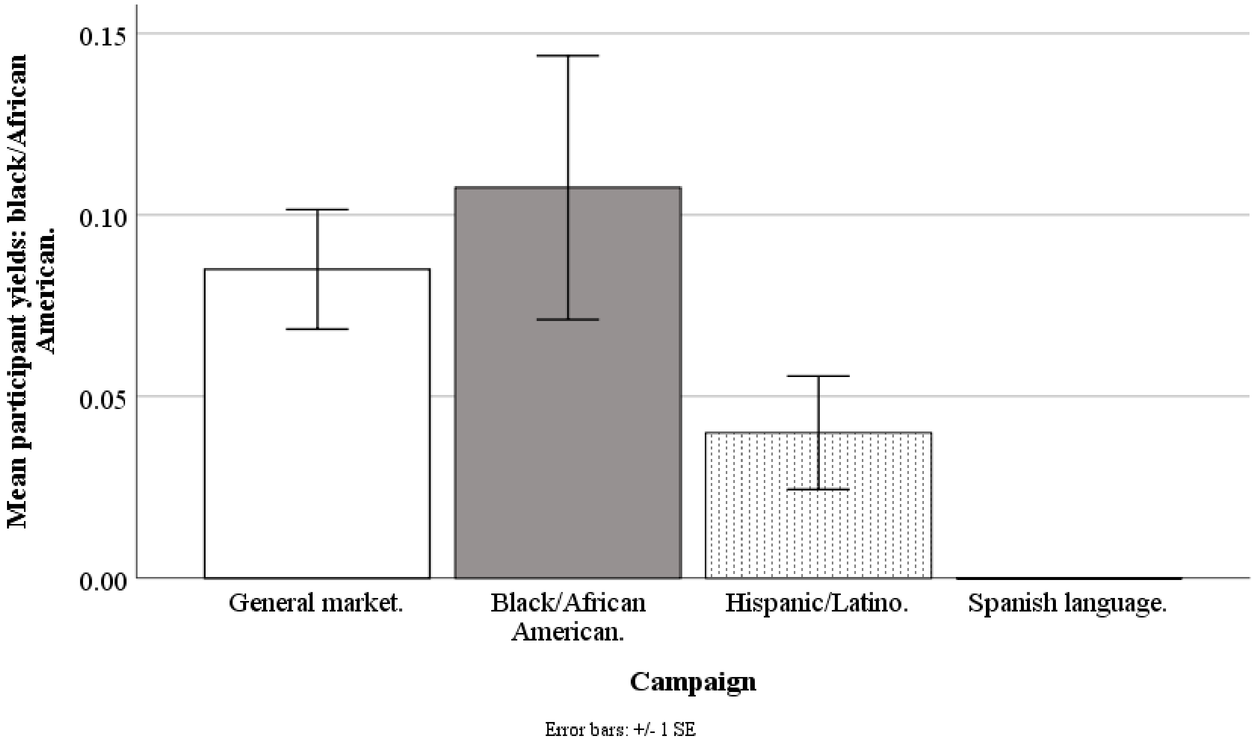
Figure 7. Mean participant yields of whites for Facebook advertising campaigns.



Black/African American Participant Yields

The Facebook campaigns marginally differed on weekly mean counts of participants who were black/African American ($F_{3,190}$=2.14; P=.10), with the general market–focused campaign at 0.09 (SD 0.13), black/African American–focused at 0.11 (SD 0.28), Hispanic-/Latino-focused at 0.04 (SD 0.12), and Spanish language–focused at .001 (SD 0.001). However, none of the pairwise comparisons between specific campaigns were significant (all P>.25; Figure 8).

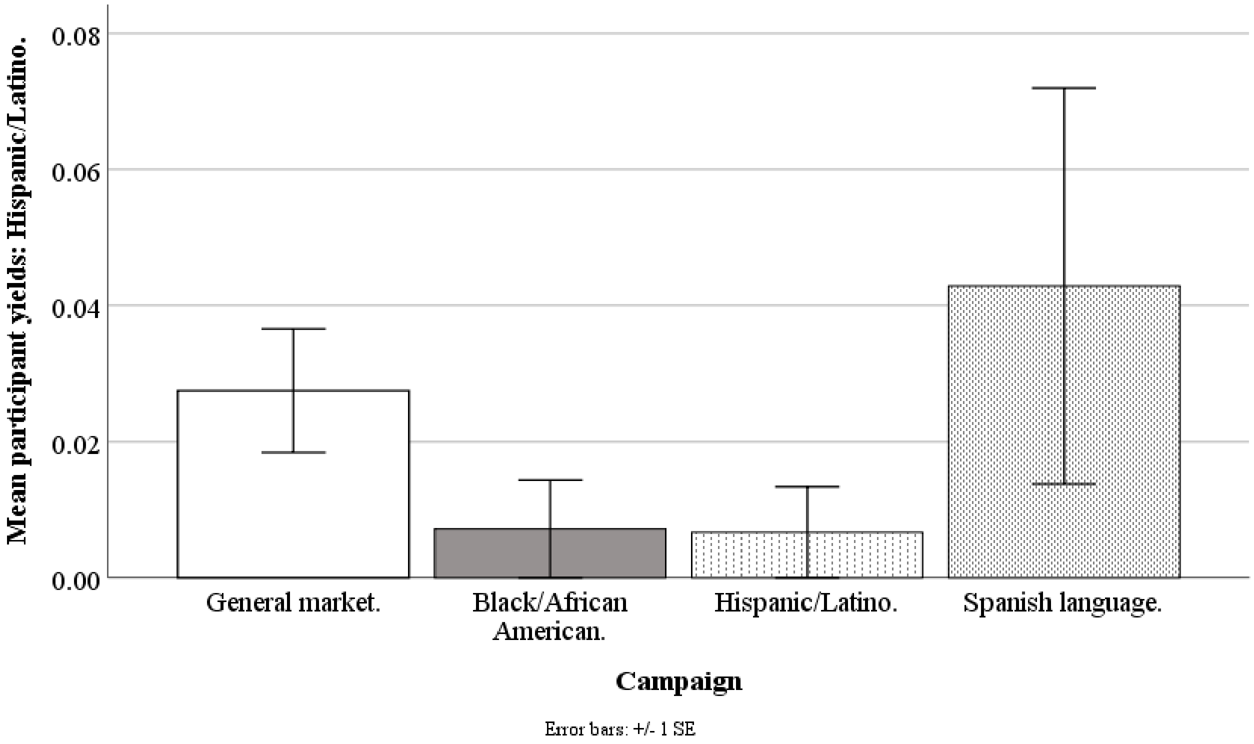
Figure 8. Mean participant yields of blacks/African Americans for Facebook advertising campaigns.



Hispanic/Latino Participant Yields

The Facebook campaigns marginally differed on the weekly mean counts of participants who were Hispanic/Latino ($F_{3,190}$=2.25; P=.08), with the general market–focused campaign at 0.03 (SD 0.07), black-/African American–focused at 0.01 (SD 0.06), Hispanic-/Latino-focused at 0.01 (SD 0.05), and Spanish language–focused at 0.04 (SD 0.11). However, none of the pairwise comparisons between specific campaigns were significant (all P>.31; Figure 9).

Figure 9. Mean participant yields of Hispanics/Latinos for Facebook advertising campaigns.



Results Regarding Costs by Ethnic Group

Overview of Cost Analysis

Facebook advertising costs are available at the campaign level (for instance, for a Hispanic-/Latino-focused campaign) but not at the individual level (eg, for a Hispanic/Latino individual) because at the individual level, internal auctions are used to determine if an advertisement is shown, and Facebook does not share these individual-level costs with its advertisers [32]. The Facebook Ads Manager only provides costs per campaign and per day, and so our best estimate of the recruitment cost of any individual regardless of ethnicity was the average cost for the campaign on the day in which the recruitment took place. Daily data were aggregated to weekly data for accuracy and smoothing (eg, staff did not work weekends to convert queries to applicants). Our cost analysis is based on weekly data from June 20, 2017, to September 17, 2018, excluding major holidays (n=61 weeks), except the Spanish language data that are from June 20, 2017, to September 30, 2018 (n=15 weeks). Refer to Figures 10-13 for campaign costs by ethnicity at the advertisement click, query, applicant, and participant stages.

Figure 10. Mean costs per advertisement click for Facebook advertising campaigns.

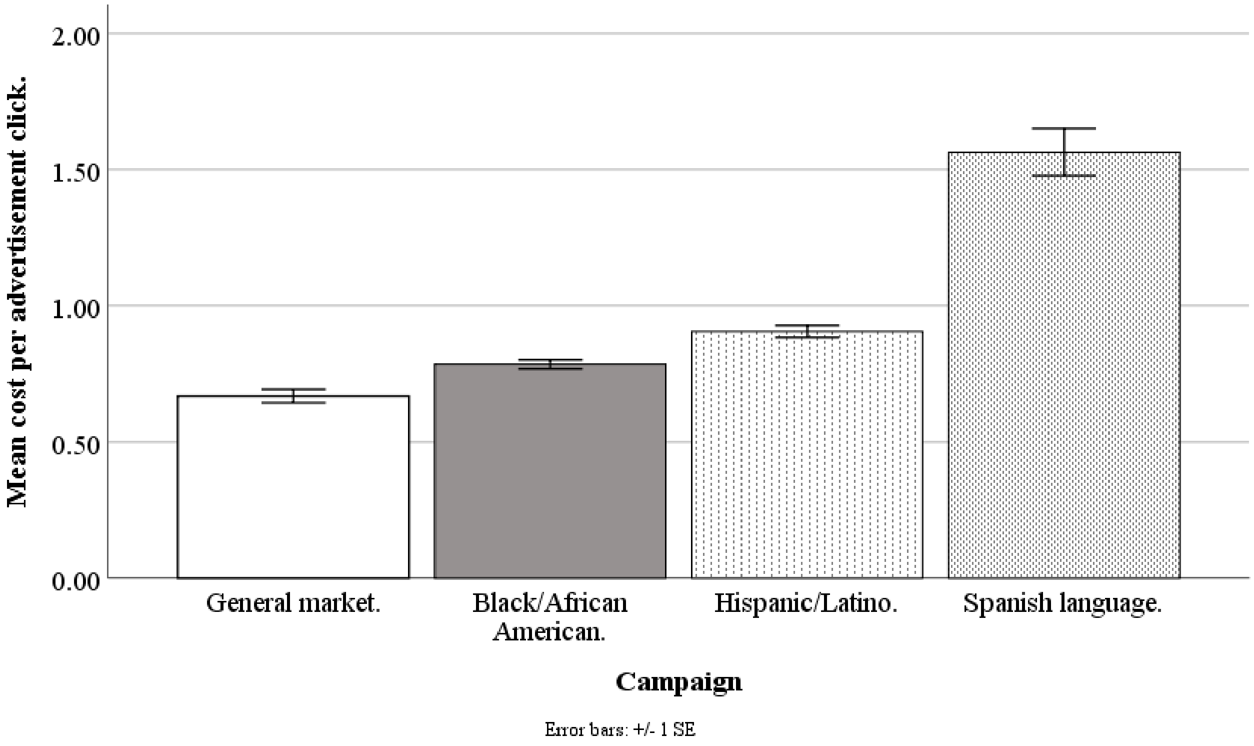


Figure 11. Mean costs per query for Facebook advertising campaigns.

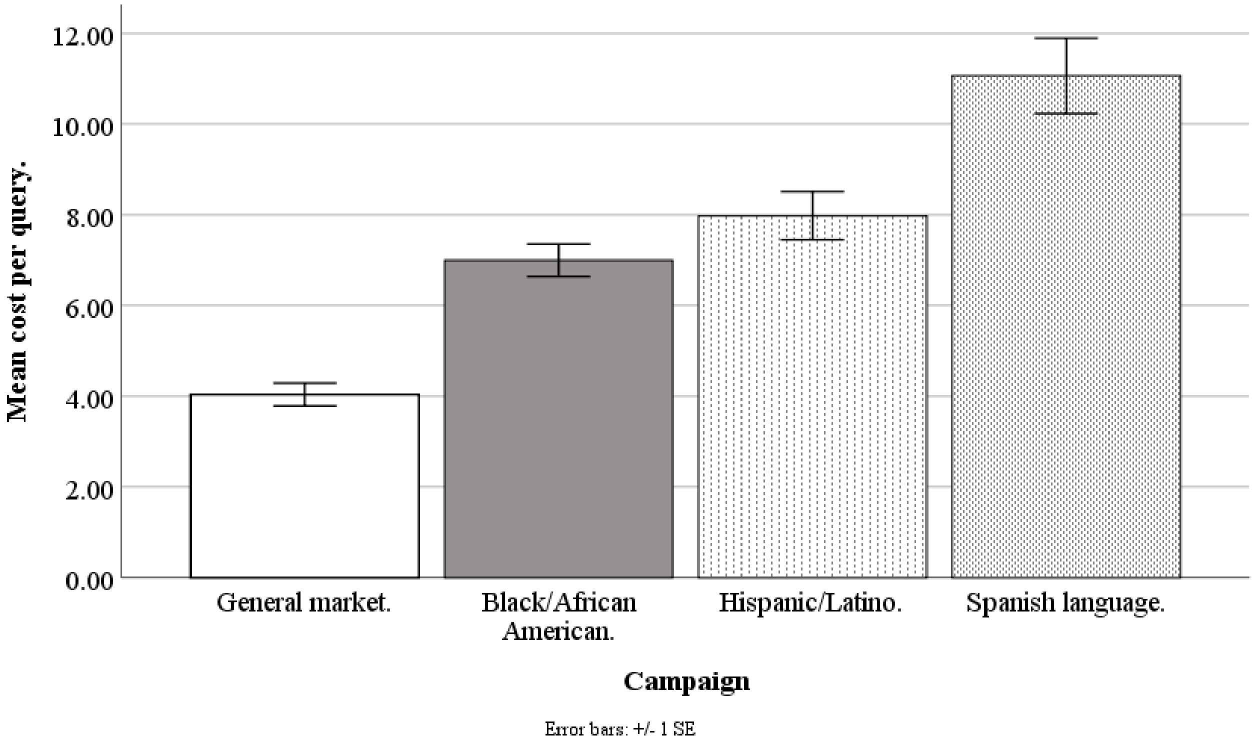


Figure 12. Mean costs per applicant for Facebook advertising campaigns.

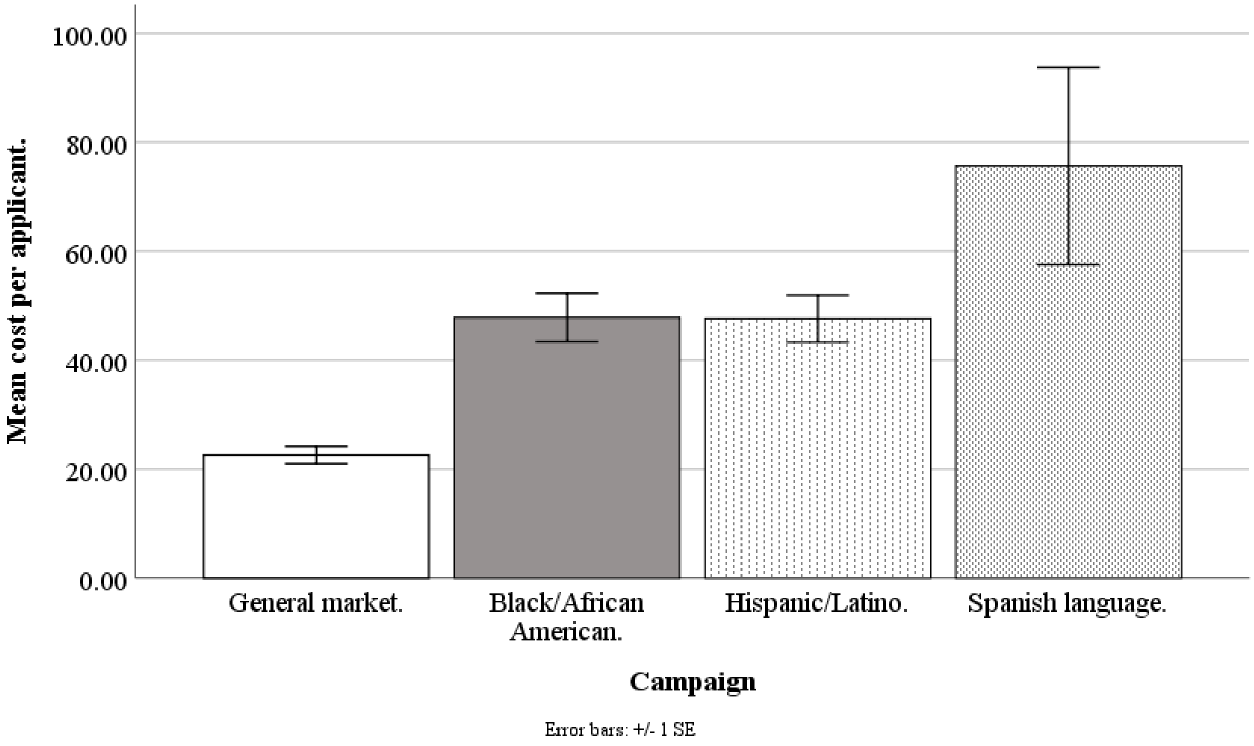
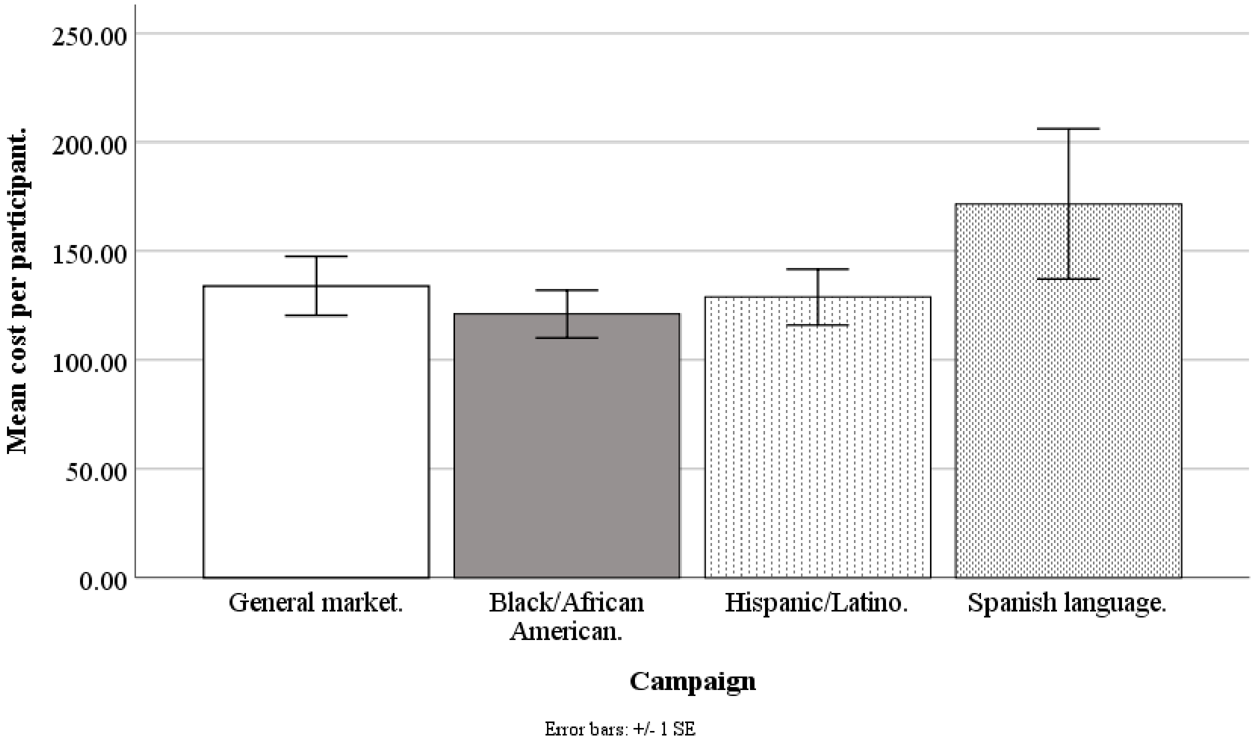


Figure 13. Mean costs per participant for Facebook advertising campaigns.



Cost per Advertisement Click

The Facebook campaigns differed on cost per advertisement click ($F_{3,194}$=97.18; P<.001), with the general market–focused campaign at US $0.67 (SD 0.19), black-/African American–focused at US $0.78 (SD 0.13), Hispanic-/Latino-focused at US $0.91 (SD 0.17), and Spanish language-focused at US $1.56 (SD 0.34). The cost per advertisement click was lower for the general market–focused campaign, compared with black-/African American–focused ($t_{1,194}$=3.41; P=.004), Hispanic-/Latino-focused ($t_{1,194}$=6.97; P<.001), or Spanish language–focused ($t_{1,194}$=16.57; P<.001). The cost per advertisement click was lower for the black-/African American–focused campaign than for the Hispanic-/Latino-focused ($t_{1,194}$=3.56; P=.002) or Spanish language-focused ($t_{1,194}$=14.43; P<.001) campaigns, and lower for the Hispanic-/Latino-focused than for the Spanish language–focused campaign ($t_{1,194}$=12.49; P<.001; Figure 10).

Cost per Query

The Facebook campaigns differed in cost per query ($F_{3,194}$=28.42; P<.001), with the general market–focused campaign at US $4.04 (SD 1.99), black-/African American–focused at US $7.00 (SD 2.77), Hispanic-/Latino-focused at US $7.98 (SD 4.13), and Spanish language–focused at US $11.06 (SD 3.23). The cost per query was lower for the general market–focused campaign than for the black-/African American–focused ($t_{1,194}$=5.26; P<.001), Hispanic-/Latino-focused ($t_{1,194}$=7.01; P<.001), or Spanish language–focused campaigns ($t_{1,194}$=7.84; P<.001). The cost per query was lower for the black-/African American–focused campaign than for the Hispanic-/Latino-focused ($t_{1,194}$=3.56; P=.002) or Spanish language–focused campaigns ($t_{1,194}$=4.54; P<.001), and lower for the Hispanic-/Latino-focused than for the Spanish language–focused campaign ($t_{1,194}$=3.45; P=.004; Figure 11).

Cost per Applicant

The Facebook campaigns differed in cost per applicant ($F_{3,182}$=13.80; P<.001), with the general market–focused at US $22.61 (SD 12.21), black-/African American–focused at US $47.82 (SD 33.00), Hispanic-/Latino-focused at US $47.63 (SD 32.16), and Spanish language–focused at US $75.65 (SD 67.60). The cost per applicant was lower for the general market–focused campaign, compared with the black-/African American–focused ($t_{1,182}$=4.26; P<.001), Hispanic-/Latino-focused ($t_{1,182}$=4.23; P<.001), or Spanish language–focused campaigns ($t_{1,182}$=5.61; P<.001). The cost per applicant was comparable for the black-/African American–focused campaign than for the Hispanic-/Latino-focused campaign ($t_{1,182}$=0.03; P=.99) but lower than that of the Spanish language–focused campaign ($t_{1,182}$=2.92; P=.02). The cost per applicant was also lower for the Hispanic-/Latino-focused than for the Spanish language–focused campaign ($t_{1,182}$=2.94; P=.02; Figure 12).

Cost per Participant

The Facebook campaigns were comparable on cost per participant ($F_{3,121}$=0.73; P=.54), with the general market–focused at US $133.90 (SD 103.92), black-/African American–focused at US $121.05 (SD 60.02), Hispanic-/Latino-focused at US $128.90 (SD 67.86), and Spanish language–focused at US $171.66 (SD 97.37). Likewise, comparing the specific campaigns pairwise, there were no differences in cost per participant (all P>.62; Figure 13).

Results of Cambridge Analytica Scandal

Overview of Analysis of Scandal Effects

In March 2018, Facebook started to receive negative publicity because a third-party company, Cambridge Analytica, had allegedly harvested information from more than 50 million Facebook accounts without users' permission [33]. Our advertisements ran before and after this scandal broke, and so we compared our costs for advertisement clicks, queries, applicants, and participants across these 2 time periods, to determine whether the Facebook privacy scandal increased our recruitment costs, and whether this depended on the campaign: general market–focused, black-/African American–focused, or Hispanic-/Latino-focused (our Spanish language campaign was no longer running). Our scandal analysis was based on weekly data from June 20, 2017, to September 17, 2018, excluding major holidays (n=61 weeks). Because the original New York Times article on the scandal appeared on March 17, 2018, all weeks before March 20, 2018, (n=33) were treated as prescandal and the remaining weeks (n=28) were treated as postscandal. Refer to Figures 14-17 for our costs by campaign pre- versus postscandal.

Figure 14. Mean costs per advertisement click for Facebook advertising campaigns before versus after privacy scandal.

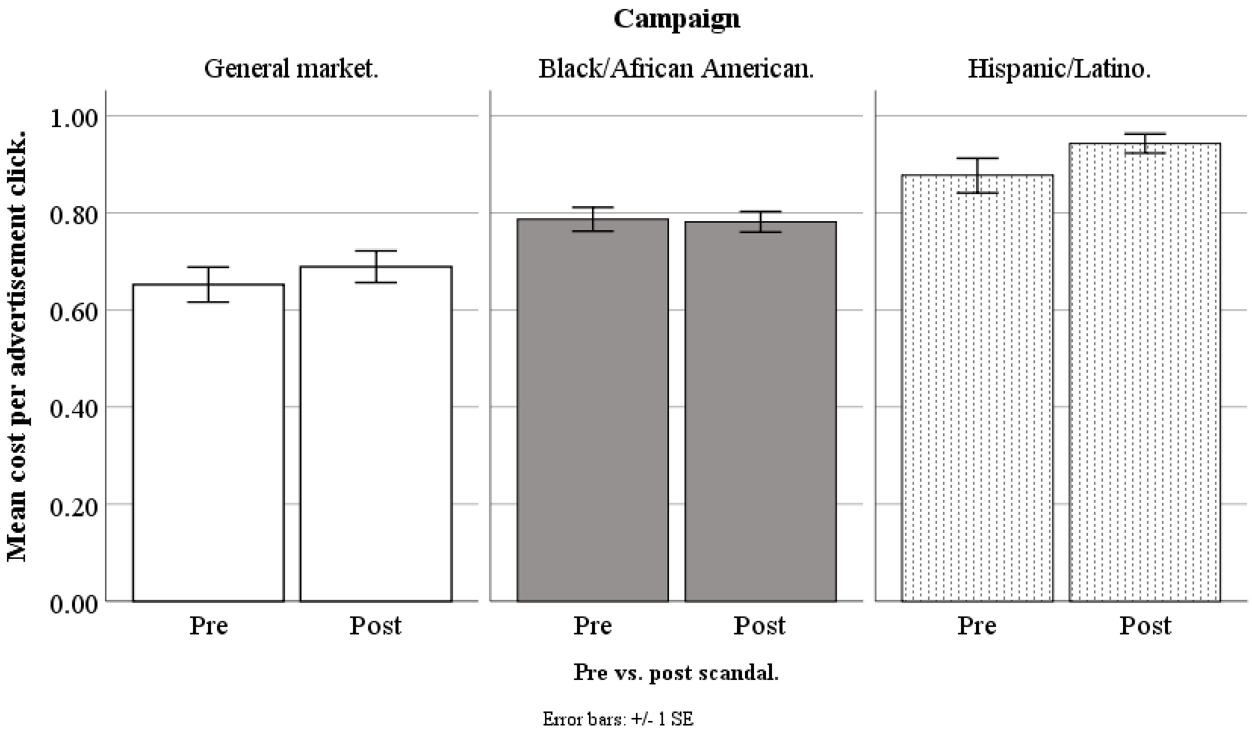


Figure 15. Mean costs per query for Facebook advertising campaigns before versus after privacy scandal.

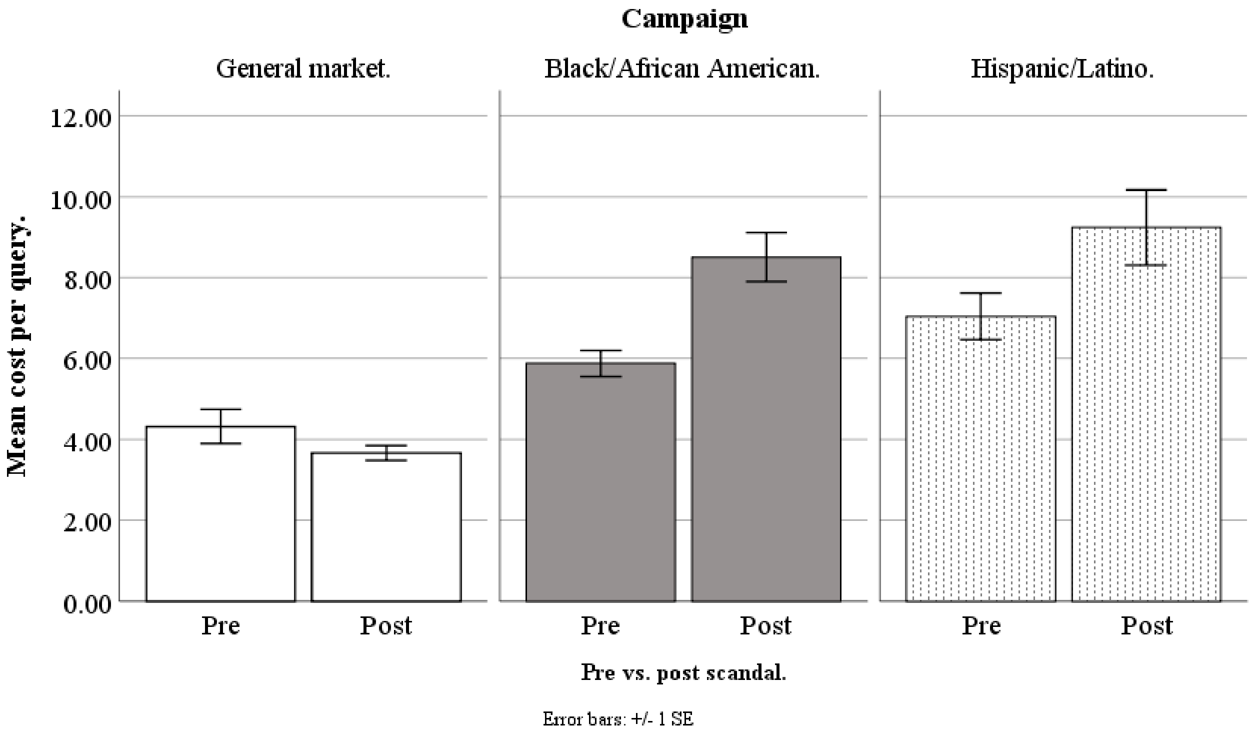


Figure 16. Mean costs per applicant for Facebook advertising campaigns before versus after privacy scandal.

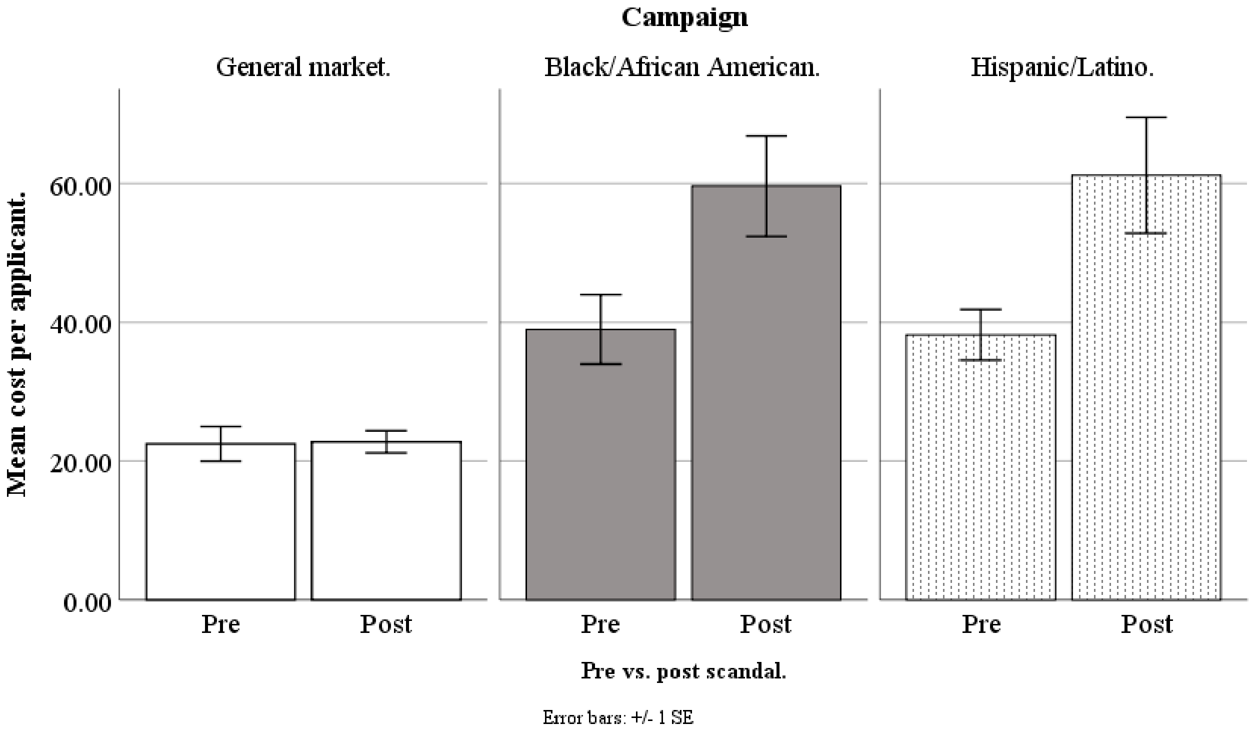
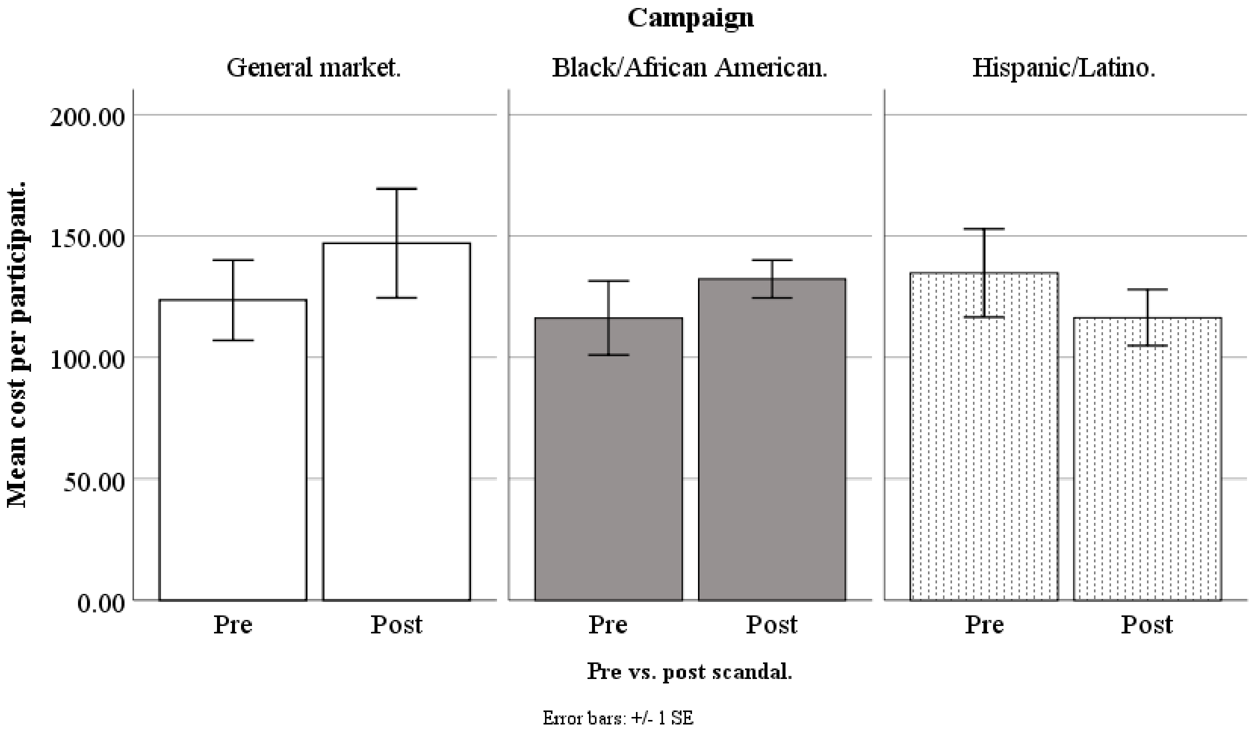


Figure 17. Mean costs per participant for Facebook advertising campaigns before versus after privacy scandal.



Cost per Advertisement Click

On cost per advertisement click, the effect for pre- versus postscandal was not significant ($F_{1,177}$=1.67; P=.20). Additionally, there were no significant differences in cost per advertisement click when we conducted specific pairwise comparisons of the campaigns (all P>.13; Figure 14).

Cost per Query

On cost per query, there was a significant effect for pre- versus postscandal ($F_{1,177}$=9.93; P=.002) and a scandal by campaign type 2-way interaction ($F_{1,177}$=5.46; P=.005), meaning the scandal effect varied by campaign. For the general market–focused campaign, there was no scandal effect on cost per query ($F_{1,177}$=0.74; P=.39), with prescandal at US $4.32

(SD 2.49) and postscandal at US $3.66 (SD 0.94). For the black-/African American–focused campaign, there was a scandal effect on cost per query ($F_{1,177}$=11.84; *P*=.001), with prescandal at US $5.87 (SD 1.89) and postscandal higher at US $8.51 (SD 3.08). Likewise, for the Hispanic-/Latino-focused campaign, there was a scandal effect on cost per query ($F_{1,177}$=8.28; *P*=.01), with prescandal at US $7.04 (SD 3.39) and postscandal higher at US $9.24 (SD 4.74; Figure 15).

Cost per Applicant

On cost per applicant, there was a significant effect for pre- versus postscandal ($F_{1,166}$=13.29; *P*<.001) and a scandal by campaign type 2-way interaction ($F_{1,166}$=3.30; *P*=.04), meaning the scandal effect varied by campaign. For the general market–focused campaign, there was no scandal effect on cost per applicant ($F_{1,166}$=.002; *P*=.97), with prescandal at US $22.48 (SD 14.68) and postscandal at US $22.77 (SD 8.19). For the black-/African American–focused campaign, there was a scandal effect on cost per applicant ($F_{1,166}$=8.65; *P*=.004), with prescandal at US $38.96 (SD 28.31) and postscandal higher at US $59.64 (SD 35.63). Similarly, for the Hispanic-/Latino-focused campaign, there was a scandal effect on cost per applicant ($F_{1,166}$=10.57; *P*=.001), with prescandal at US $38.19 (SD 21.20) and postscandal higher at US $61.19 (SD 40.08; Figure 16).

Cost per Participant

On cost per participant, the effect for pre- versus postscandal was not significant ($F_{1,111}$=0.15; *P*=.70). Additionally, there were no significant differences in cost per participant when we conducted specific pairwise comparisons of the campaigns (all *P*>.31; Figure 17).

Results Regarding Lookalike Audience

We used Facebook's lookalike audience feature for a short time to determine if it reduced our recruitment costs relative to our baseline general market–focused campaign. Facebook's lookalike audience feature involves a proprietary method of targeting individuals similar to a chosen source audience, that is, an audience that responded favorably to a past campaign based on a specific website response as recorded by the Facebook Pixel. On the basis of the designated source audience, Facebook seeks to generate a similarly interested audience [34]. We designated our source audience as individuals who filled out our query form on our study website, which was recorded by our website's Facebook Pixel. Our analysis was based on daily data from January 3, 2018, to March 26, 2018 (n=26 days). It compared the Facebook lookalike campaign that sought to optimize our queries to our standard general market–focused campaign, which ran simultaneously. We assessed cost per advertisement click, query, and applicant; however, cost per participant could not be estimated due to the short time window. This test was ended soon after it began because of privacy concerns. Facebook's lookalike audience feature has the drawbacks of an unknown and unknowable algorithm and unclear use of Facebook user data to create the algorithm, which might possibly involve using Facebook users' network of friends. Therefore, we could not fully describe the data collection approach to our potential study participants and ask for their informed consent or obtain their friends' consent, if needed.

The cost per advertisement click for the lookalike audience campaign was lower than the general market–focused campaign ($F_{1,50}$=26.64; *P*<.001), with the lookalike audience at US $0.73 (SD 0.30) and general market–focused at US $1.07 (SD 0.14). However, the cost per query for the lookalike audience campaign did not differ from the general market campaign ($F_{1,49}$=0.23; *P*=.64), with lookalike audience at US $3.98 (SD 2.24) and general market at US $4.27 (SD 2.19). The cost per applicant for the lookalike audience campaign also did not differ from the general market campaign ($F_{1,40}$=2.75; *P*=.11), with lookalike audience at US $15.58 (SD 6.07) and general market at US $20.13 (SD 10.13).

Results Regarding Audience Network Placement

Audience network placement is a feature recommended by Facebook to its advertisers, which involves placing their advertisements on other websites or mobile apps rather than on Facebook itself, to reach a broader audience. We assessed whether using Facebook's audience network placement reduced recruitment costs. We tried out the audience network placement very early on, before we began study recruitment, and so only costs per advertisement click and query were available. Facebook automatically provides advertisement click data, and we were able to record queries because we had installed and monitored our query form on our website. We used the audience network placement on its own during 2 test weeks in 2016, so we could attribute all queries to it. Later on, our Facebook Pixel would identify queries by campaign.

Here, we compare data from our general market campaign that used audience network placement (August 3, 2016-August 10, 2016, and August 23, 2016-August 30, 2016, n=16 days) to standard advertisement placement on Facebook alone (1 year later, on exactly the same dates, n=16 days). The analysis shows that the cost per advertisement click was lower when the audience network placement was used compared with placement on Facebook alone ($F_{1,30}$=262.10; *P*<.001), with audience network at US $0.11 (SD 0.05) and Facebook alone at US $0.69 (SD 0.13). However, the cost per query was substantially higher when audience network placement was used compared with placement on Facebook alone ($F_{1,30}$=4.26; *P*=.048), with the audience network at US $13.40 (SD 14.83) and Facebook alone at US $5.62 (SD 2.76). After 2 test weeks, based on noticeably poor query results, we stopped using the audience network placement.

Discussion

Principal Findings on Zip Code Targeting

Our results indicate that researchers can successfully recruit diverse individuals for web-based health-related studies using Facebook advertising campaigns with zip code targeting. By placing advertisements in zip codes ranked first on the percentage of households of the target ethnicity and then on the focal health behavior (in this case, cigarette expenditure per household), we successfully increased our outreach to and yield of black/African American– and Hispanic/Latino applicants. Table 1 shows our expected yields if we had continued these campaigns for 2 years, spending US $20 per day or US $140 per week, compared with our baseline general market campaign. Although we were concerned about higher costs, costs were not significantly higher for these ethnic-focused campaigns compared with our baseline general market campaign. We also recruited Hispanic/Latinos using Facebook's option to reach Spanish language speakers, and this produced a similar yield to our Hispanic/Latino campaign using zip codes. But it was more costly to produce this yield, perhaps because our study required English fluency.

Table 1. Estimated campaign yields based on spending US $20/day on each campaign for 2 years.

Yield type and campaign[a]	White, n (%)	Black/African American, n (%)	Hispanic/Latino, n (%)	Total[b], N
Queries				
General market	2005 (84.5)	294 (12.4)	74 (3.1)	2373
Black/African American	1125 (51.9)	986 (45.5)	58 (2.7)	2169
Hispanic/Latino	1258 (66.9)	301 (16.0)	322 (17.1)	1881
Spanish language	730 (62.0)	62 (5.3)	385 (32.7)	1177
Applicants				
General market	428 (86.6)	56 (11.3)	10 (2.0)	494
Black/African American	269 (68.1)	115 (29.1)	11 (2.8)	395
Hispanic/Latino	249 (77.3)	36 (11.2)	37 (11.5)	322
Spanish language	185 (72.5)	16 (6.3)	54 (21.2)	255
Participants				
General market	83 (87)	9 (9)	3 (3)	95
Black/African American	28 (70)	11 (27)	1 (2)	40
Hispanic/Latino	22 (81)	4 (14)	1 (3)	27
Spanish language	29 (87)	0 (0)	4 (12)	33

[a] The estimates are based on the yield of US $140/week (US $20/day) multiplied by 104 weeks to show estimated results for 2 years. The estimates in this table come from the observed yields of these campaigns extrapolated to reflect a standard expenditure of US $20/day for 2 years. The observed yields are based on US $77,133 spent, out of the US $87,425 in total spending on Facebook for the randomized controlled trial. The remainder was spent before the Facebook Pixel install or on weeks or test campaigns not reported here. The US $77,133 spent was divided up across campaigns as follows: general market 427 days at US $126/day, black/African American 427 days at US $23/day, Hispanic/Latino 427 days at US $23/day, and Spanish language 105 days at US $36/day.

[b] Row percentages add up to 100. Other ethnicities that were recruited during the campaigns were not factored into these estimates.

Principal Findings on Audience Network Placement and Scandal Effects

This study also found that Facebook's recommended audience network placement, although cheaper at producing advertisement clicks, was far more expensive at producing website queries. This means that many individuals who clicked on our advertisements were not interested. Many individuals may have clicked on our advertisements to use a free app (eg, play a game), not because they were interested in our study. Thus, we do not recommend using the audience network placement. We also learned that the Facebook privacy scandal involving Cambridge Analytica had a negative impact on recruitment for our campaigns that were black/African American–focused and Hispanic-/Latino-focused. This scandal increased our cost per query and cost per applicant. Hence, researchers should keep Facebook scandals in mind, as they may raise recruitment costs.

Principal Findings on Audience Lookalike Feature

Facebook's recommended audience lookalike feature produced advertisement clicks at a lower price point than our baseline general market campaign; however, query and applicant price points were similar to our general market campaign. Researchers may want to consider using Facebook's audience lookalike feature because it may cost less at the advertisement click stage, if this is a high priority, and if the required *source audience* can be identified based on website behavior as tracked by a Facebook Pixel (we used queries on our website). Potential drawbacks to this feature include privacy and institutional review board (IRB) concerns. Facebook's audience lookalike feature seems to involve using data collected whenever people are logged into Facebook, including data that are both public and relatively more private, including likes, posts, visits to other websites, friend networks, and friends' web-based behavior. This data use could raise potential privacy and IRB approval issues because it is not possible in IRB consent forms or information sheets to describe to potential participants what Facebook data of theirs will be used for the audience lookalike feature (this is proprietary) and it is not possible to obtain consent from their Facebook friend networks.

Principal Findings on Advertising Costs

Facebook-based recruitment for health-related research has been the focus of previous studies, allowing for cost and yield comparisons with our study. Looking at cost per advertisement click, Whitaker et al [21] reported a mean of US $0.57 across several studies (range US $0.20-US $1.74), whereas Ramo and Prochaska reported a mean of US $0.45 [35]. Our costs per advertisement click across campaigns are within this range (US $0.67, SD 0.19-US $1.56, SD 0.34). Cost per participant varies widely; Whitaker et al [21] found that the cost per participant ranged from US $1.36 to US $110.00 depending on study length and engagement. Although our cost per participant ranged from US $121.05 (SD 60.02) to US $171.66 (SD 97.37), we recruited for a 3-month study that involved daily engagement in a web-based group and complete cessation from smoking. We were able to generate detailed cost and yield findings because we directed our website developer to install a Facebook Pixel on our study website. This Pixel allowed us to track which Facebook advertisement campaign a query came from and determine whether the individual who queried applied to the study and participated. We, therefore, recommend that researchers install the Facebook Pixel.

Study Limitations and Strengths

Our findings on recruitment yields standardize for spending at US $140 per week, so they are relatively independent of costs. However, our findings on recruitment costs are limited by the fact that recruitment costs are strongly affected by study inclusion/exclusion criteria, participation time demands, and study benefits and incentives. Recruitment costs can also fluctuate daily and seasonally. In addition, we recruited throughout the continental United States for our web-based study, and studies that are more local may yield different outcomes. Moreover, our costs per yield for a Spanish language campaign may be overestimated because English fluency was a requirement for inclusion in our study. Studies with translated materials and no English-speaking requirements may find higher enrollments and lower costs.

A strength of our study is that it was a national study with a large sample size; 980 people were recruited overall. Furthermore, we examined Facebook advertising campaigns considering both yields and costs and considered 4 separate outcomes: advertisement clicks, queries, applicants, and participants. In addition, our study lasted for 61 weeks, allowing us to compare costs of Facebook recruitment both before and after a major Facebook privacy scandal. Our main goal was to innovatively test the efficacy of targeting specific zip codes on Facebook to reach ethnically or racially diverse populations. Advertiser intent can have ethical ramifications when utilizing zip code targeting, especially when protected characteristics are targeted such as race or ethnicity. As of August 2019, Facebook restricts audience targeting options, including zip code targeting in the case of housing, credit, or employment to help prevent discrimination [36]. In our case, advertiser intent is heath promoting, and therefore zip code targeting is justifiable.

Conclusions

Our main conclusion is that Facebook advertising campaigns that employ suitable zip code targeting can help to find and recruit blacks/African Americans and Hispanics/Latinos for web-based health-related studies.

Acknowledgments

This research was funded by the National Cancer Institute Grant # R01CA204356-04 and the Stanford Cancer Institute Grant # P30CA1244351. The sponsors were not involved in any of the studies or analyses.

Conflicts of Interest

Unrelated to this project, JP has provided consultation to pharmaceutical and technology companies that make medications and other treatments for quitting smoking and has received funding from Facebook for planning evaluation of a mobile health intervention. JP and CP have served as expert witnesses in lawsuits against tobacco companies.

Multimedia Appendix 1
Facebook advertising campaign timeline.
[DOCX File , 22 KB - jmir_v22i6e17554_app1.docx]

Multimedia Appendix 2
CONSORT diagrams for participants by Facebook campaign.
[DOCX File , 125 KB - jmir_v22i6e17554_app2.docx]

References

1. Armstrong K, Putt M, Halbert CH, Grande D, Schwartz JS, Liao K, et al. Prior experiences of racial discrimination and racial differences in health care system distrust. Med Care 2013 Feb;51(2):144-150 [FREE Full text] [doi: 10.1097/MLR.0b013e31827310a1] [Medline: 23222499]
2. Jacobs E, Mendenhall E, Mcalearney AS, Rolle I, Whitaker E, Warnecke R, et al. An exploratory study of how trust in health care institutions varies across African American, Hispanic and white populations. Commun Med 2011;8(1):89-98. [doi: 10.1558/cam.v8i1.89] [Medline: 22616359]

3. Kennedy BR, Mathis CC, Woods AK. African Americans and their distrust of the health care system: healthcare for diverse populations. J Cult Divers 2007;14(2):56-60. [Medline: 19175244]
4. Allmark P. Should research samples reflect the diversity of the population? J Med Ethics 2004 Apr;30(2):185-189 [FREE Full text] [doi: 10.1136/jme.2003.004374] [Medline: 15082815]
5. Oh SS, Galanter J, Thakur N, Pino-Yanes M, Barcelo NE, White MJ, et al. Diversity in clinical and biomedical research: a promise yet to be fulfilled. PLoS Med 2015 Dec;12(12):e1001918 [FREE Full text] [doi: 10.1371/journal.pmed.1001918] [Medline: 26671224]
6. Reid RJ, Peterson NA, Lowe JB, Hughey J. Tobacco outlet density and smoking prevalence: does racial concentration matter? Drugs 2009 Jul 10;12(3):233-238. [doi: 10.1080/09687630500035485]
7. Sautter E, Oretskin NA. Tobacco targeting: the ethical complexity of marketing to minorities. J Bus Ethics 1997;16:1011-1017. [doi: 10.1023/A:1017909702926]
8. Stoddard JL, Johnson CA, Sussman S, Dent C, Boley-Cruz T. Tailoring outdoor tobacco advertising to minorities in Los Angeles county. J Health Commun 1998;3(2):137-146. [doi: 10.1080/108107398127427] [Medline: 10977250]
9. Soneji S, Ambrose BK, Lee W, Sargent J, Tanski S. Direct-to-consumer tobacco marketing and its association with tobacco use among adolescents and young adults. J Adolesc Health 2014 Aug;55(2):209-215 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.01.019] [Medline: 24661738]
10. Timberlake DS, Pechmann C. Trends in the use and advertising of discount versus premium snuff. Nicotine Tob Res 2013 Feb;15(2):474-481. [doi: 10.1093/ntr/nts160] [Medline: 23024245]
11. Luke D, Esmundo E, Bloom Y. Smoke signs: patterns of tobacco billboard advertising in a metropolitan region. Tob Control 2000 Mar;9(1):16-23 [FREE Full text] [doi: 10.1136/tc.9.1.16] [Medline: 10691754]
12. Hafez N, Ling PM. Finding the Kool Mixx: how Brown & Williamson used music marketing to sell cigarettes. Tob Control 2006 Oct;15(5):359-366 [FREE Full text] [doi: 10.1136/tc.2005.014258] [Medline: 16998169]
13. Yerger VB, Przewoznik J, Malone RE. Racialized geography, corporate activity, and health disparities: tobacco industry targeting of inner cities. J Health Care Poor Underserved 2007 Nov;18(4 Suppl):10-38. [doi: 10.1353/hpu.2007.0120] [Medline: 18065850]
14. Kool Market Development Program. 1083. Unknown Kool Market Development Program URL: https://www.industrydocuments.ucsf.edu/docs/#id=lqcy0025 [accessed 2020-04-30]
15. Clement J. Statista. 2019. Number of Facebook users in the United States from 2017 to 2023 URL: https://www.statista.com/statistics/408971/number-of-us-facebook-users/ [accessed 2019-10-20]
16. Perrin A, Anderson M. Pew Research Center. 2019. Share of US Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018 URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/?utm_source=sailthru&utm_medium=email&utm_term=ABN_MorningMediaNewsfeed&utm_campaign=MorningMediaNewsfeed_Newsletter_2019041108&s_id=5c954803a777392e08683e66 [accessed 2019-10-19]
17. Krogstad JM. Pew Research Center. 2015. Social Media Preferences Vary by Race and Ethnicity URL: https://www.pewresearch.org/fact-tank/2015/02/03/social-media-preferences-vary-by-race-and-ethnicity/ [accessed 2020-04-13]
18. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
19. Carter-Harris L, Bartlett Ellis R, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. J Med Internet Res 2016 Jun 15;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
20. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 2014 Feb;16(2):247-251. [doi: 10.1093/ntr/ntt157] [Medline: 24127266]
21. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
22. Geller SE, Koch A, Pellettieri B, Carnes M. Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress? J Womens Health (Larchmt) 2011 Mar;20(3):315-320 [FREE Full text] [doi: 10.1089/jwh.2010.2469] [Medline: 21351877]
23. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27:1-28. [doi: 10.1146/annurev.publhealth.27.021405.102113] [Medline: 16533107]
24. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. Tob Control 2017 Mar;26(2):188-194 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052768] [Medline: 26928205]
25. Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ. Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. J Med Internet Res 2015 Feb 23;17(2):e50 [FREE Full text] [doi: 10.2196/jmir.3772] [Medline: 25707037]

26. Bala D. Facebook for Developers. 2017. News for Developers: Targeting Exclusions-Update URL: https://developers.facebook.com/ads/blog/post/v2/2017/12/19/targeting-exclusions-update-blog-post/ [accessed 2020-02-04]
27. US Census Bureau. 2017. 2013-2017 ACS 5-year Estimates URL: https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2017/5-year.html [accessed 2020-04-28]
28. Experian. 2019. ConsumerView: Marketing Data That Connects Brands With Fans URL: https://www.experian.com/marketing-services/targeting/data-driven-marketing/consumer-view-data.html [accessed 2019-10-20]
29. US Bureau of Labor Statistics Latest Numbers. 2019. BLS Data Finder URL: https://www.bls.gov/bls/data_finder.htm [accessed 2020-04-28]
30. US Bureau of Labor Statistics Latest Numbers. 2019. Consumer Expenditure Surveys URL: https://www.bls.gov/cex/csxgeography.htm [accessed 2019-10-18]
31. Facebook for Business. 2019. About Facebook Pixel URL: https://www.facebook.com/business/help/651294705016616?helpref=uf_permalink [accessed 2019-10-22]
32. Facebook for Business. 2019. About Ad Auctions URL: https://www.facebook.com/business/help/430291176997542?id=561906377587030 [accessed 2019-10-20]
33. Rosenberg M, Confessore N, Cadwalladr C. New York Times. 2018. How Trump Consultants Exploited the Facebook Data of Millions URL: https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html [accessed 2020-04-28]
34. Facebook for Business. 2019. About Lookalike Audiences URL: https://www.facebook.com/business/help/164749007013531?helpref=uf_permalink [accessed 2019-10-19]
35. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012 Feb 23;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
36. Facebook for Business. 2019. Updates To Housing, Employment and Credit Ads in Ads Manager URL: https://www.facebook.com/business/news/updates-to-housing-employment-and-credit-ads-in-ads-manager [accessed 2020-02-26]

Abbreviations

IRB: institutional review board

Edited by G Eysenbach; submitted 19.12.19; peer-reviewed by C Zheng, J Folk; comments to author 20.01.20; revised version received 13.02.20; accepted 29.03.20; published 05.06.20.

Please cite as:
Pechmann C, Phillips C, Calder D, Prochaska JJ
Facebook Recruitment Using Zip Codes to Improve Diversity in Health Research: Longitudinal Observational Study
J Med Internet Res 2020;22(6):e17554
URL: https://www.jmir.org/2020/6/e17554
doi:10.2196/17554
PMID:32501274

Original Paper

Spelling Errors and Shouting Capitalization Lead to Additive Penalties to Trustworthiness of Online Health Information: Randomized Experiment With Laypersons

Harry J Witchel[1], PhD; Georgina A Thompson[1], BMBS; Christopher I Jones[2], CSci, CStat, PhD; Carina E I Westling[3], PhD; Juan Romero[4]; Alessia Nicotra[4], MBBS, PhD, FRCP; Bruno Maag[4], TGmeF, HFG; Hugo D Critchley[1], MBChB, DPhil, FRCPsych

[1]Department of Neuroscience, Brighton and Sussex Medical School, Brighton, United Kingdom
[2]Department of Primary Care and Public Health, Brighton and Sussex Medical School, Brighton, United Kingdom
[3]Faculty of Media and Communication, Bournemouth University, Bournemouth, United Kingdom
[4]Dalton Maag Ltd, London, United Kingdom

Corresponding Author:
Harry J Witchel, PhD
Department of Neuroscience
Brighton and Sussex Medical School
Trafford Centre for Medical Research
Brighton
United Kingdom
Phone: 44 1273 873 549
Email: h.witchel@bsms.ac.uk

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/4/e29452

Abstract

Background: The written format and literacy competence of screen-based texts can interfere with the perceived trustworthiness of health information in online forums, independent of the semantic content. Unlike in professional content, the format in unmoderated forums can regularly hint at incivility, perceived as deliberate rudeness or casual disregard toward the reader, for example, through spelling errors and unnecessary emphatic capitalization of whole words (online *shouting*).

Objective: This study aimed to quantify the comparative effects of spelling errors and inappropriate capitalization on ratings of trustworthiness independently of lay insight and to determine whether these changes act synergistically or additively on the ratings.

Methods: In web-based experiments, 301 UK-recruited participants rated 36 randomized short stimulus excerpts (in the format of information from an unmoderated health forum about multiple sclerosis) for trustworthiness using a semantic differential slider. A total of 9 control excerpts were compared with matching error-containing excerpts. Each matching error-containing excerpt included 5 instances of misspelling, or 5 instances of inappropriate capitalization (*shouting*), or a combination of 5 misspelling plus 5 inappropriate capitalization errors. Data were analyzed in a linear mixed effects model.

Results: The mean trustworthiness ratings of the control excerpts ranged from 32.59 to 62.31 (rating scale 0-100). Compared with the control excerpts, excerpts containing only misspellings were rated as being 8.86 points less trustworthy, those containing inappropriate capitalization were rated as 6.41 points less trustworthy, and those containing the combination of misspelling and capitalization were rated as 14.33 points less trustworthy (*P*<.001 for all). Misspelling and inappropriate capitalization show an additive effect.

Conclusions: Distinct indicators of incivility independently and additively penalize the perceived trustworthiness of online text independently of lay insight, eliciting a medium effect size.

(J Med Internet Res 2020;22(6):e15171) doi:10.2196/15171

KEYWORDS

communication; health communication; persuasive communication; online social networking; trust; trustworthiness; credibility; typographical errors

Introduction

Trustworthiness of Online Health Information: Background, Context, and Importance

As of 2019, 90% of all British adults use the internet at least weekly [1], and as patients, they often search online for health information to solve their medical problems [2]; furthermore, they are likely to be influenced by the online information, including changing their health care decisions and their frequency of ambulatory care visits [3]. In a survey of 2200 adults with chronic health conditions in the United States who were active social media users, 57% used a health condition–specific website (eg, specializing in multiple sclerosis or rheumatoid arthritis) on a monthly basis, and 5% used such sites daily; half of the patients surveyed had asked a health-related question to others online within the previous 6 months, and 87% of those were seeking responses from other patients with the health condition [4].

In response to this explosion of unvetted potential sources providing online health care information that is acted upon, researchers, experts, and medical professionals have repeatedly expressed concern about the inaccuracy of online information and the limited ability of lay consumers to adequately assess its validity [2,5,6]. In particular, it has long been known that when lay users determine whether to use and trust online health care information, they are strongly influenced by nonmedical criteria that experts do not use [7-11]. Broadly, although academics favor a checklist approach of transparency criteria [12], the approach of nonexperts appears to be more variable and situation dependent, and it prioritizes factors such as ease of understanding and the attractiveness of graphic design [9]. More generally, in research on the factors that influence judgments of trust, all aspects of trustworthiness can be relational and depend on the type of person (or group being studied); major relational factors include accessibility, both cognitive as well as physical [13], and correctly accommodating language to the intended audience [14].

Understanding lay assessment of the trustworthiness of online health information is important because false online information presented to the general public, if believed, has the potential to undermine correct medical advice [15], to elicit unhealthy behavior [16], and to influence sociopolitical discourses on health care and other topics [17-19]. Sbaffi and Rowley’s [20] recent review of how laypeople assess the trustworthiness and credibility of online health information concluded that so far, much less research has been performed to understand what interferes with trust (we describe these as penalties to trustworthiness [21]), compared with what causes trust.

Trust and credibility are closely related to one another and information quality, although there remains disagreement among researchers as to the exact definitions of these terms. In general, the definitions emphasize the likelihood of information use, believability, reliability, and dependability [20,22]. Correspondingly, there is little evidence to suggest that the general population reliably make fine conceptual distinctions between trustworthiness and credibility of information.

Online health support often is divided into seeking information or emotional reassurance and can be gender specific (eg, prostate cancer vs breast cancer) and person specific [23,24]. Trustworthiness remains pertinent to online emotional reassurance and sharing, as shown by the many occurrences of large-scale hoaxes designed to manipulate emotions [8]. For example, between 2010 and 2011, a Macmillan cancer forum was inundated with posts in response to an elaborate hoax by a purported mother about her 6-year-old daughter struggling with cancer. On the exposure of the hoax (perpetrated by a lonely 16-year-old girl), many users of the forum—who had formed close online relationships with the supposed mother—refused to believe it had all been a hoax [25]. In light of the range of uses that online health information has for the general public, we contend that it is important to build knowledge of the factors that influence how trust and its absence are formed online beyond source analysis and fact checking. As this study shows, both linguistic and metalinguistic factors affect how trustworthiness is instinctively rated.

Theoretical Underpinnings for Factors Influencing Trustworthiness

The range of elements that influence online trustworthiness includes issues related to the sources (eg, author identification and the absence of advertising), issues related to the content (eg, a date stamp and inclusion of medical evidence), issues related to design and engagement (eg, inclusion of images), and issues affecting all the above (eg, the absence of typographical errors) [20,26]. For some time, credibility has been subdivided into aspects such as source credibility, message credibility, and media credibility, which strongly influence each other [22]. All such aspects of credibility include accessibility, which can be relational and cognitive as well as physical [13]. More recently, Sun et al [6] have divided the elements that influence consumer evaluation of online health information into 25 *criteria* and 165 *quality indicators*; criteria are rules that reflect notions of value and worth (eg, expertise and objectivity), whereas quality indicators are properties of information objects to which criteria are applied to form judgments (eg, the owner of the website and inclusion of statistics). In line with Diviani et al [5], Sun et al [6] have suggested that indicators can be positive or negative (in terms of trustworthiness), and that consumers’ perceived online health information quality could conceivably be measured by a small set of core dimensions (ie, a few groups of criteria might incorporate many of the quality indicators that explain most of the trustworthiness judgments).

So far, Lederman et al [8] have proposed a five-category model that highlights verification processes via the comparison of different websites or other online statements. This model includes argument quality, source credibility, source literacy competence, and crowd consensus. An extended six-category model [26] proposes that the lay reader may assess some or all of the following: reputation, endorsement, consistency with other sources, self-confirmation (agreement with the reader's own opinions), persuasive intent, and expectancy violation. For this research, we have adopted Lederman et al's [8] model of credibility; however, the choice among these models for this research is moot because all the models [5,6,20,22,26] are concordant with the idea that spelling errors will detract from judgments of trustworthiness (ie, they are negative quality indicators).

Incivility, Literacy Competence, and Errors of Writing Mechanics

Although institutionally produced websites and curated online health content (cf. [6]) will usually be both grammatically correct and civil, the responses by the general public may be uncivil [27], and patient-authored text is known to have many misspellings [28] (and so do discharge summaries written by doctors [29]). Inappropriate capitalization (including inappropriate capitalization of entire words, sometimes called online *shouting*) and misspelling are examples of errors in literacy competence [8] and writing mechanics [30]; writing mechanics is defined as elements of a language that only manifest when communication is in written form. Both inappropriate capitalization and misspelling have been highlighted by qualitative investigations as explicit criteria used by lay readers in judgments of online credibility [8,31]. The rationales given to explain why these two error types undermine trustworthiness are that either (1) the errors imply a lack of intelligence (expertise, ability, authority, and education) [8,32,33] or (2) that they suggest a lack of motivation (objectivity, attention to detail, and conscientiousness) to be trustworthy [34,35]. The term incivility is used to describe this latter lack of motivation or effort to make statements that are compliant with rules of communication. Hargittai [33] distinguishes between spelling errors (errors because of the levels of education or social inequalities) and typographical errors (errors because of accidentally hitting the wrong keys on the keyboard). The fact that these errors are not corrected during proofreading is another form of incivility. Incivility implies a lack of respect for the reader, the platform, and the rules of social exchange [36,37], and it is fundamentally relational. Quantitative research on incivility in the mainstream World Wide Web demonstrates that civil statements are rated as more trustworthy and influential than uncivil ones [38,39].

Integration Versus Heuristics: Lay Judgments Based on Multiple Cues

The reader who must judge unvetted online health information is faced with a wealth of cues that indicate the degree of trustworthiness, and those cues may have contradictory effects (eg, a cogent message that is misspelled). There are three broad theories for how individuals (both lay and expert) make judgments based on multiple cues. The rational approaches are represented by the information integration model [40], in which an individual accounts for all the different pieces of information by a complex (but often subconscious) mathematical process that is usually based on addition, multiplication, or averaging; extensive observations of integration in judgments occur across cultures and individuals. This is the process that Sun et al [6] allude to when they propose that the consumer integrates the relevant trustworthiness criteria and quality indicators in a "complex cost-benefit analysis." In 2003, Fogg's [41] Prominence-Interpretation Theory proposed explicit mathematical relationships for how to predict the effects of multiple factors on credibility, but the theory never detailed how to measure the relevant quantities independently. Computational models typically assume that the elements of incivility (eg, inappropriate capitalization) act together either additively or nonlinearly [42-44], although hypothesis-led proof for this assumption is minimal. The elucidation of this integration is only just starting in the literature [45], and the relative importance of each indicator in this intuitive cost-benefit analysis remains unknown; the relative values for each cue may be elucidated empirically by statistical methods such as regression, but it is unlikely that these values would be explicit or transparent in the minds of most lay decision makers. This is an area that needs to be further researched.

An alternative theory to cost-benefit analysis (ie, for how judgments are made based on multiple cues) is the process-level cognitive perspective [46]; this has been made famous by the heuristics and biases research program from behavioral economics [47]. Heuristics are rapid cognitive strategies or shortcuts (either explicit or subconscious) formulated as practical, bounded rational decision systems for multiple cues that can be more transparent than complex cost-benefit analyses, for example, hierarchical lexicographic decision models [48]. In a take-the-best judgment [49], first, a single cue (the most important one) is searched for in the environment and considered independently of all others, and if a tie or no clear result occurs, then the second most important cue is sought and decided upon, and so on. For example, when deciding whether you have the right of way when driving your car through an intersection, first, you follow the signal of any policemen present, and only if there are no police present, do you seek and consider a traffic light (including a temporary traffic light for construction), and only if there are no traffic lights present, do you then consider static road markings and the positions of the other cars.

In biased heuristics, only a limited subset of information (often only one cue) is used to make a fast and ecological judgment outside of conscious awareness [46]; when biased, these heuristics are used to support preferred or preconceived outcomes. With biased heuristics, the prioritization of the cue, and even the cue's basic validity for the judgment being made, is dubious. Such biased heuristics have been used to explain seemingly irrational preferences that individuals make in situations involving slot machines and organ donation [50]. Examples of biased heuristic processes include representativeness, where some cues are weighted disproportionately compared with their real representativeness—and availability—where a cue that is easily recalled (such as occurs with recency and news) determines the judgment. None of these judgment models so far proposed have explicitly assessed specific issues within literacy competence.

Determining the Criteria and Weighting of Judgments Based on Multiple Cues

A key feature of heuristics is that they are typically made subconsciously, and that post hoc explanations for such choices are often self-serving justifications or rationalizations [51]. That is, in the case of heuristics, the decision maker does not have privileged information on how the judgment was made, and furthermore, the decision maker can be wrong about themselves [52]. For example, university students have been shown to greatly overestimate how much they actually learned from excellent lecturers (conflating it with how much they feel they learned, which is discounted by effort and exertion) [53]. This creates potential issues for researchers, in which insight-based techniques (both qualitative interviews and quantitative questionnaires) can lead to judgments and explanations of causes that are inaccurate because of demand characteristics or social desirability [54,55]. It has long been known that when evaluating credibility (eg, website privacy policy), the importance of factors that lay individuals say they would use to make their judgment diverge from the factors they are observed to use [41]. Recent sophisticated measurements suggest that although multiple factors can interact when quantifying perceived credibility (of information for an online health forum), these interactions do not support Petty and Cacioppo’s Elaboration Likelihood Model (ELM) [56]. Furthermore, there is a gap in the literature for any data showing how spelling errors interact with other writing mechanics violations that might also affect trustworthiness [57]. This suggests that lay insight into the influences on their perceptions of trustworthiness are imperfect, and that research on trustworthiness should be supplemented by approaches that do not rely on such insight.

To avoid insight bias from our lay participants, the approach of this study is to compare and contrast *marginal differences* in penalties to trustworthiness elicited by different combinations of literacy or writing mechanics violations. Our experiments on changes in marginal trustworthiness were specifically organized so that the participating healthy volunteers were unaware that the experiment tested the effects of capitalization or misspelling *per se* (with full institutional ethical approval). We did not explicitly ask lay individuals for their beliefs regarding how their process for judging message credibility incorporates quality indicators relating to source credibility and media credibility. That is, instead of asking directly, “How much less would you trust a web post that is misspelled?” we simply presented the participants with some posts that had misspellings and asked the same question as usual, “How trustworthy do you find this information?” while subtly varying misspelling and capitalization (ie, *shouting text*, not acronyms or beginnings of sentences).

Research Questions

The research questions (RQs) were as follows:

- RQ1: Do errors in writing mechanics and incivility lead to marginal changes (ie, without signposting) in judgments for message trustworthiness?
- RQ2: Are the marginal effects of incivility, such as with inappropriate capitalization, on trustworthiness judgments quantitatively comparable with the known effects of writing mechanics such as spelling errors?
- RQ3: How, if at all, do the effects of writing mechanics (eg, spelling errors) and incivility (eg, inappropriate capitalization) integrate? Is there a ceiling effect or a binary effect in which once a message source is judged as incompetent, no further trustworthiness penalty is added to the judgment (similar to a take-the-best heuristic) [49]? Or is there some additive, multiplicative, synergistic, or otherwise integrative function that increases the penalty on the judgment to a new higher level when both cues are present [40]?

Methods

Participants, Recruitment and Ethical Approval

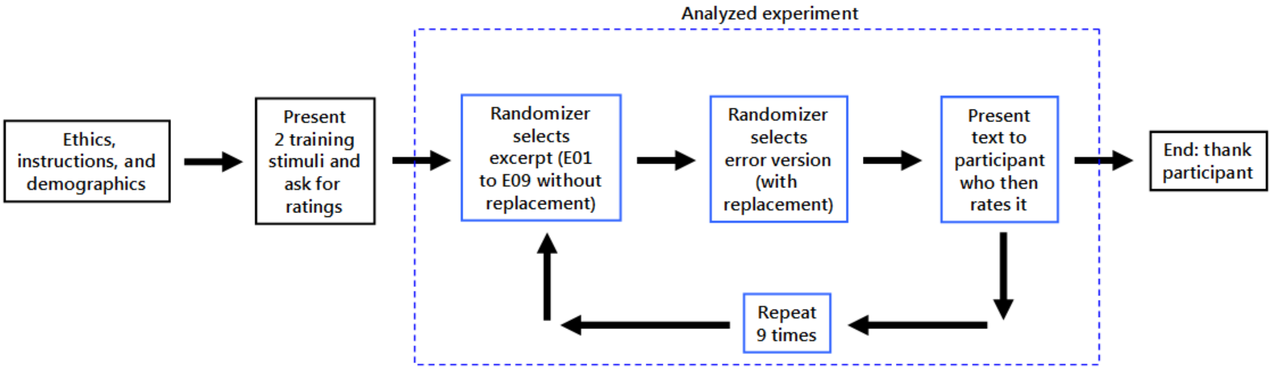
The project was approved by our local ethics committee (Brighton and Sussex Medical School’s Research Governance and Ethics Committee, University of Sussex, approval 16/044/WIT). All experiments were performed according to the Declaration of Helsinki. All individuals provided informed consent via a welcome page in each online study. Participants were recruited via the Prolific website. We specified that the recruitment should focus on UK-based members of the public. Participants consented to participate with the understanding that the research concerned *responses to online text*; none of the advertising, web URLs or experimental information to the participants mentioned that the experiment was related to graphics, formatting, incivility, and spelling. This feature of the advertising was approved by our ethics committee, not least because the paragraphs were not considered misleading or potentially emotionally adverse.

Study Design Process

The study design was a confirmatory, cross-sectional experiment with a balanced incomplete block design; it was a randomized experiment with lay participants, each of whom experienced only a limited number of the possible options. The total number of experimental excerpts that were tested in the entire cohort was 36; there were 9 excerpts, each in 4 possible versions: no errors, inappropriate capitalization only (Caps only), misspelling only, or the combination of both errors. The stimulus texts (and the versions) are shown in the Multimedia Appendix 1. There were two additional paragraphs that always appeared as the first two stimulus texts, which were training stimuli. The training stimuli were not labeled as being different from other stimuli in any way and were never included in the statistical analyses. The only purpose of the training stimuli was to allow participants to familiarize themselves with the rating task and with the range of the trustworthiness scale: training stimulus 1 was quite believable (mean trustworthiness rating 54.33, SD 23.39; n=301), and training stimulus 2 was less plausible (mean trustworthiness 40.20, SD 25.80; n=301; $P<.001$, paired t test).

Each participant experienced and rated only 9 (of the possible 36) experimental excerpts, and those 9 included exactly one version of each excerpt, and among those 9, that participant would experience a mix of different error types (Figure 1). For example, participant number 001 experienced the capitalization-only version of excerpt E07 and then the *both* *errors* (capitalization and misspelling) version of excerpt E02. This double randomization prevented any participant from seeing the same excerpt twice.

Figure 1. Experiment design: what each participant experiences.



The goal was to present to participants a set of coherent text excerpts of similar lengths (70-100 words) on the topic of multiple sclerosis, each excerpt being a coherent answer to a question. The rationale for presenting excerpts in a question-then-answer format was that it was possible to ask whether readers trusted the advice enough to believe it or act upon it. To make these excerpts, a range of comments were found in the public domain (Table 1); these texts often had to be edited substantially to fit within the word count or to avoid explicitly endorsing commercial products (see Multimedia Appendix 2 for a comparison of the original and presented texts).

Initially, the excerpts and the presentation system were tested online by a small group of testers, who then provided verbal feedback on the test to the experimental team. After that, a cohort of 40 participants was recruited online to test the paragraphs and demonstrate that the paragraphs elicited similar standard deviations in ratings (19-28 units out of 100) and elicited a wide range of mean trustworthiness ratings (from 20 to 70). These excerpts were to be presented either as they were (*no errors* or the negative control) or in one of the three error versions stated above.

Table 1. Sources of excerpts on multiple sclerosis.

Code	Brief topic description	Website	Length
L01	Numerous artificial sweeteners	blogspot.com	78
L02	Hoax about artificial sweeteners	quora.com	81
E01	Triggers of the immune system	healingchronicles.com	81
E02	Programmer’s intelligence	dailystrength.org	89
E03	Epstein Barr virus	medicaldaily.com	92
E04	Avonex patient	quora.com	74
E05	Up there in risk	quora.com	96
E06	Mental exercises	dailystrength.org	85
E07	Vitamin D	quora.com	100
E08	Small risk of Progressive Multifocal Leukoencephalopathy	my-ms.org	90
E09	Half of all people	ms.pitt.edu	71

Text Interventions

For the error versions of the texts (inappropriate capitalization or misspelling), we wanted to include 5 of the relevant errors per excerpt (a total of 10 errors for the combination of both errors version), with the errors spread throughout the excerpt (rather than bunched together). In excerpts where inappropriate capitalization was required, there would be 5 *sets* of words or phrases. Normally, a set was 1 or 2 words, although one of the 5 sets had to be a 4-word series. The priorities for selecting words to capitalize were (in this order) as follows:

- Adverbs (especially those suggesting extremity such as *very* or *never*)
- Judgments (*rubbish*, *hopeless*, and *horror*)
- Strong emotions (*worried* and *angry*)
- Words implying danger (*fatal* and *death*)
- Amounts (*all*, *every*, and *ten*)
- Adjectives (rather than nouns)
- Action verbs (especially gerunds)
- Conjunctions (*and*)

To verify that each word that we capitalized was naturally capitalized on the web, we analyzed words that were capitalized online on Twitter. We used the Claritin corpus, which is a crowdsourced data set of all the Twitter tweets that contained the word *claritin* in the month of October 2012 [58]. This corpus has some 4900 tweets, and we used MATLAB (MathWorks) to find all the words that were in all capitals (which did not have a hashtag or an at-sign in them); this resulted in a list of 343 capitalized words (Multimedia Appendix 3), many of which were short words, acronyms, and internet memes. From this list of words spontaneously capitalized on the web, we selected words in our excerpts to capitalize.

The rationale for how we selected words to misspell was that misspellings should be quite noticeable, and that the meaning of the words should remain clear to the reader even when misspelled. We avoided homonyms and words that looked plausibly English when misspelled. To make sure that misspelled words were noticeable, short words were preferred, or we placed the misspellings in the first syllable of a multisyllable word. In addition, one of the misspelled words had to be in the first 5 words of a paragraph. The misspelled word had to be completely understandable (in the absence of other words or context) even when misspelt. Thus, a misspelled word with missing or added letters should be pronounceable in English (eg, *yu* plainly means *you*). The types of misspellings were as follows:

- Swap one letter for another letter that is next to it on a qwerty keyboard (*pisitive*)
- Double a consonant (*esstimate*)
- Double a vowel or add an extra vowel (*theere*)
- Leave out a vowel (*expsure*)

To verify that each word that we misspelled was naturally misspelled on the web, we searched for the misspelled word along with the words *health* and *forum*; if we could not find at least two examples of a misspelling on online health forums, we did not use it. A complete listing of the misspellings and where we found them online is in Multimedia Appendix 4.

Study Delivery

The study was presented to participants using the Qualtrics portal, which allows for a wide range of question types and keeps track of answers and total response time. A full description of the survey in the Checklist for Reporting Results of Internet E-Surveys format [59] is included as part of the Multimedia Appendix 5 for this paper. The web-based study welcome page explained in brief what the study was about and what it entailed (estimated 8 min participation time, including reading the ethics and filling in demographics), the ethics of the study (include the ability to withdraw at any time), and a brief complaints procedure. The ethics page explicitly excluded participants aged younger than 18 years or those from vulnerable populations. A pointer to a full-length participant information sheet (3 pages) was shown; the welcome/ethics page had an “I agree” button at the bottom. After the welcome page, participants filled in a brief multiple-choice demographics page, which included questions on gender, age, field of work (eg, health care, agriculture, and retired), and familiarity with the English language/Roman alphabet. All demographic questions included an option for “rather not say.” After the demographics page, participants saw the instructions page and then were launched into the experimental ratings pages.

Each ratings page consisted of a short excerpt of text (which was randomized as to whether or not it had the spelling or capitalization errors), followed by a horizontal slider for rating how *trustworthy* the participant found the statement to be; the slider had anchors of *completely untrustworthy* (Figure 2, left) and *completely trustworthy* (Figure 2, right). Although the data collection was numerical (0-100, left to right), there were no numerical cues or tic marks seen by the participants. The instructions to the participants for the trustworthiness ratings were, “If you find something trustworthy, you would be prepared to act upon it; an untrustworthy statement you would ignore, and a rating in the middle represents information where you would want more proof or confirmation that it is correct.” As explained in the instructions to participants, each stimulus excerpt was written as if it was an answer to a question written on an unmoderated health forum, with a specific focus on multiple sclerosis. Multiple sclerosis was chosen as a topic because the information was obviously important, but healthy participants would be unfamiliar with the veracity of each statement; thus, we predicted that the trustworthiness ratings would be more susceptible to nonverbal cues. The questions were as follows: (1) Is multiple sclerosis preventable? (2) How risky is Tecfidera as a treatment for multiple sclerosis? and (3) Does multiple sclerosis decrease intelligence/IQ? The nominal responses to these questions were the experimental stimulus excerpts being rated.

Figure 2. Unnumbered horizontal slider for trustworthiness ratings.

Study Design, Analysis, and Statistics

Sample Size

To detect a difference in rating scores between two groups (ie, no errors vs misspellings, inappropriate capitalization, or both), with 80% power at the 5% significance level, assuming the standard deviation within each group is 25 and the difference between groups is 15 (equivalent to a medium effect size of 0.517), would require 60 participants per group (120 total). Each participant was asked to rate 9 text excerpts, randomly divided between the 4 conditions, and assuming an intraclass (ie, within participant) correlation coefficient of 0.185 gives a design effect of 2.48. The product of the design effect and the sample size for a nonrepeating experiment is 2.48×120≈300 total.

Modeling

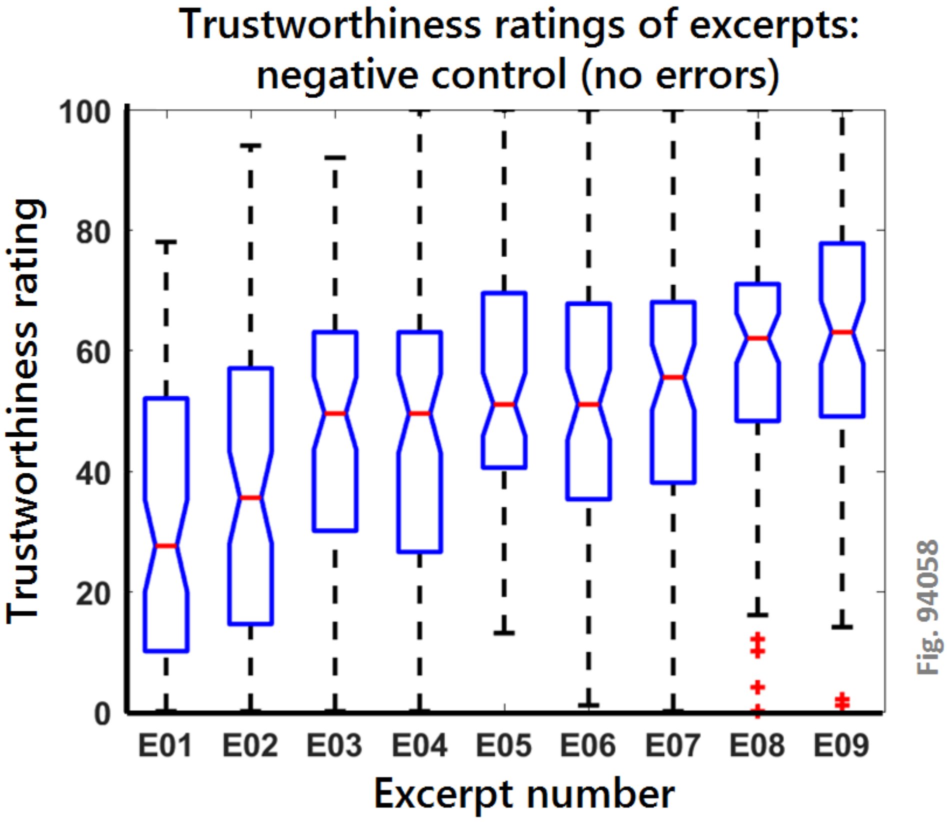
Linear mixed effects (LME) models were fitted in Stata version 16.0 using the *mixed* command. Residuals were checked for normality and homoscedasticity at the cluster and individual levels. Where the assumption of residual homoscedasticity was not appropriate, robust standard errors were used to allow for the calculation of appropriate 95% CIs and *P* values [60].

Results

Differences Between the Paragraphs

We ran an experiment in which we gathered data from 301 volunteers, each making 9 experimental observations (2709 ratings in total). We estimated that this task (rating 11 excerpts plus instructions and demographics) would take 8 min; in fact, it took a mean time of 6.25 min (mean 375.74 seconds, SE of mean 12.94 seconds). The median trustworthiness ratings of each of the excerpts (E01 to E09) in the negative control condition (ie, without any errors or incivility) are shown in the box and whisker plot in Figure 3. On each box, the red line in the center indicates the median, and the bottom and top edges of the box indicate the first and third quartiles, respectively. When the notches for two boxes do not overlap, this implies that the true medians do differ with 95% CI. The whiskers extend to the most extreme data points not considered outliers, whereas outliers are shown using red plus signs; outliers are any points that fall more than 1.5 times the IQR away from the main box. Excerpt 01 generally elicits low ratings of trustworthiness (median 27.5), whereas excerpts 08 and 09 generally elicit high ratings of trustworthiness (medians 62 and 63, respectively). As illustrated by the nonoverlapping notches, these two sets of excerpts elicit significantly different ratings of trustworthiness, which has been true in all our previous cohorts rating these excerpts for trustworthiness (data not shown). Excerpts 03, 04, 05, and 06 all elicited median ratings in the middle range from 45 to 55, whereas excerpt 02 is a transitional excerpt between E01 and the middle range and E07 is a transitional excerpt between the middle range and the most trustworthy excerpts.

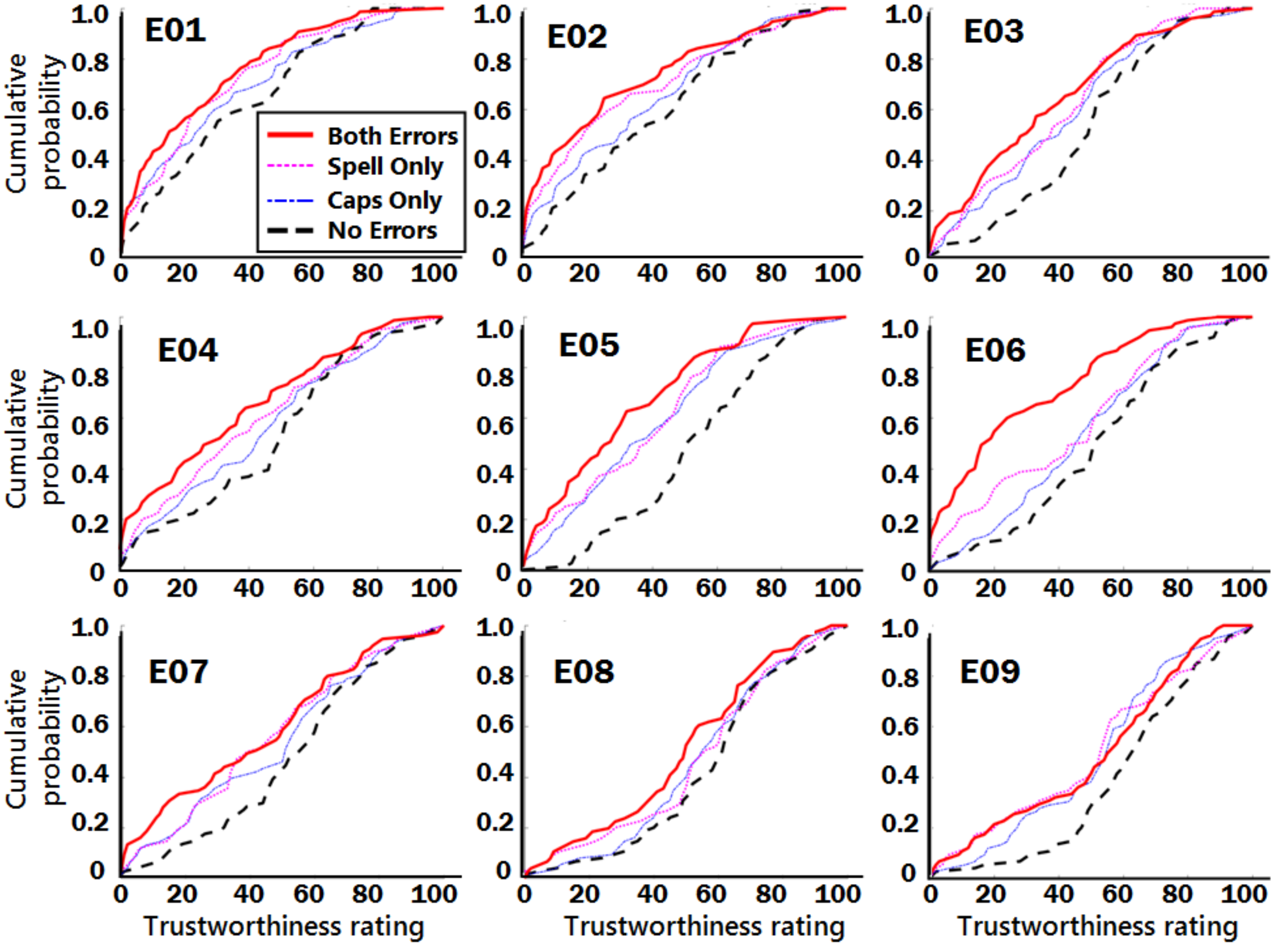
Figure 3. Trustworthiness ratings of the different excerpts in the negative control (no errors) condition. For each box, N=75.



Cumulative Distributions Shifted Left by Errors

Figure 4 shows how the errors in writing mechanics and incivility led to changes in the cumulative probability distributions for each of the excerpts. As expected, for each excerpt, compared with the negative control with no errors (black, thick dashed line), all three alterations (inappropriate capitalization (blue, thin dot-dash line), misspelling (magenta, thin dotted line), and both errors together (red, thick solid line) led to a decrease in the ratings of trustworthiness (ie, a leftward-upward shift in the curve). For most of the excerpts, at most points on the cumulative distribution curve, the combination of errors (both inappropriate capitalization and misspelling) led to decreases in trustworthiness ratings (ie, a shift left and up) compared with either of the single errors. That is, the lines for inappropriate capitalization only (blue, thin dot-dash) and misspelling only (magenta, thin dotted) fall between the thick black dashed line (no errors) and the red solid line (both errors); this result suggests that the two errors together have an additive or integrative effect.

Figure 4. Cumulative probability distribution plots for each excerpt (E01 to E09), comparing the alternative writing errors (lighter colored lines) with negative control (thick, black dashed line).



Mixed Linear Effects Model With Four Conditions of Alteration

We tested these data in a mixed linear effects model (model 1), where trustworthiness rating was the dependent variable, whereas alteration (no errors/misspellings/inappropriate capitalization/both misspellings and inappropriate capitalization) and excerpts (1-9) were included as fixed effects. A random effect for participant was included in the model to account for clustering of observations within volunteers. The reference group/condition for this model was E05, with no literacy errors. This model (and the following model) was calculated with robust standard errors [60] to allow for the heteroskedasticity of the residuals. The results of model 1 are shown in Table 2. The intracluster correlation (correlation within the individuals) coefficient estimate for model 1 is 0.334 (95% CI 0.287 to 0.384).

There is strong evidence against each of the null hypotheses that inappropriate capitalization only (result 1), misspelled only (result 2), and both errors (result 3) do not affect trustworthiness ratings compared with the negative control group; stated positively, our data suggest that there is a statistically significant penalty to trustworthiness for inappropriate capitalization (result 1), misspelling (result 2), and for both errors together (result 3). Inappropriate capitalization reduces trustworthiness ratings by −6.41 (95% CI −8.96 to −3.86), and misspelling reduces trustworthiness ratings by −8.86 (95% CI −11.61 to −6.11). The effect on trustworthiness ratings of combining both inappropriate capitalization and misspellings together is −14.33 (95% CI −17.11 to −11.55), which appears to be an additive effect.

Our further analysis aimed to test whether there was likely to be either an additive or integrative effect [40] of combining inappropriate capitalization and misspelling. Such an effect should lead to a significantly larger trustworthiness penalty when both error types are combined compared with either error individually. To test for this, an alternative specification of the LME model of the same data was formulated (model 2). In model 2, the condition *both errors* was specified as an interaction between 2 binary variables, (1) misspelling and (2) capitalization errors, instead of considering all the errors as 4 conditions in a categorical independent variable. The specification of the rest of the model was identical to model 1. In this model (Table 3), the main effects and an interaction between them provide no evidence for an interaction effect between the 2 variables; that is, the main effects for inappropriate capitalization only and for misspelled only were as in the original model 1 (Table 2), whereas the coefficient for the interaction was not significantly different from zero. This supports the interpretation that the effects of the two error types are additive, rather than partially summative or synergistic.

Table 2. Mixed effect model 1 for trustworthiness rating, with errors and excerpts as fixed effects and a random effect for the clustering of data by participant.

Categorical variables	Coefficient	Robust SE	*P* value	95% CI
Alteration				
Caps only	−6.41	1.30	<.001	−8.96 to −3.86
Misspelled	−8.86	1.40	<.001	−11.61 to −6.11
Both errors	−14.33	1.42	<.001	−17.11 to −11.55
Excerpt				
E01	−11.94	1.71	<.001	−15.28 to −8.59
E02	−8.29	1.58	<.001	−11.38 to −5.19
E03	−0.70	1.56	.65	−3.77 to 2.36
E04	0.03	1.54	.98	−2.99 to 3.05
E06	1.80	1.62	.27	−1.39 to 4.98
E07	6.19	1.71	<.001	2.83 to 9.55
E08	14.75	1.53	<.001	11.75 to 17.75
E09	14.22	1.46	<.001	11.36 to 17.09
Constant	47.06	1.53	<.001	44.06 to 50.06

Table 3. Model 2: Alternatively specified linear mixed effect model with 2 binary variables for capitalization and spelling errors and an interaction term to account for combining both types of error (all unlisted values are identical to above).

Categorical variables	Coefficient	Robust SE	*P* value	95% CI
Alteration				
Caps only	−6.41	1.30	<.001	−8.96 to −3.86
Misspelled	−8.86	1.40	<.001	−11.61 to −6.11
Interaction	0.94	1.68	.58	−2.34 to 4.24

The output for model 1 shows the following comparisons: (1) capitalization only versus no errors, (2) misspelled only versus no errors, and (3) both errors versus no errors. Table 4 shows the additional comparisons between (4) capitalization only versus misspelled only, (5) both errors versus misspelled only, and (6) capitalization only versus both errors.

Comparison 4 indicates there is weak evidence (*P*=.06) against the null hypothesis of no difference between the effects of capitalization only and misspelled only. That is, misspelled only leads to a larger trustworthiness penalty by −2.45 (95% CI −5.02 to 0.12) compared with capitalization only. There is strong evidence against the null hypothesis of there being no difference between the effects of capitalization only versus both errors combined (comparison 6). Compared with capitalization only, the combination of both errors significantly reduces trustworthiness ratings by a further −7.92 (95% CI −10.28 to −5.56). Similarly, there is strong evidence against the null hypothesis of there being no difference between the effects of misspelled only and the combination of both errors (comparison 5). Compared with misspelling only, the combination of both errors leads to a further penalty to trustworthiness ratings of −5.47 (95% CI: −7.83 to −3.11).

Table 4. All possible alterations tested in between-group comparisons (model 1).

Comparison	Coefficient	Robust SE	*P* value	95% CI
(1) Caps only versus no errors	−6.41	1.30	<.001	−8.96 to −3.86
(2) Misspelled versus no errors	−8.86	1.40	<.001	−11.61 to −6.11
(3) Both errors versus no errors	−14.33	1.42	<.001	−17.11 to −11.55
(4) Caps only versus misspelled	2.45	1.31	.06	−0.12 to 5.02
(5) Both errors versus misspelled	−5.47	1.20	<.001	−7.83 to −3.11
(6) Caps only versus both errors	7.92	1.20	<.001	5.56 to 10.28

Statistically Significant Differences Between the Excerpts

In model 1, when compared with E05, the effect of the various excerpts' contents on trustworthiness ratings varies from −11.93 to 14.75 (a range of 26.68). This range is roughly twice as large as the effect of both errors (−14.33), suggesting that the errors in incivility and writing mechanics that we tested can together have an overall effect of nearly half of the effects of the content of the excerpts we tested.

Discussion

Original Contributions

This study sought to quantitatively determine how two different errors of writing mechanics (contributing to incivility) combine to penalize subjective ratings of trustworthiness in the medically relevant context of materials typical of an unmoderated online health forum. Using an LME model of a suitably powered study, we found that all three interventions (inappropriate capitalization, misspelling, and the combination of the two) were significantly different from the negative control (no added errors or incivility), which clearly answers RQ1. The data also show that (for these 70- to 100-word long excerpts about multiple sclerosis), the trustworthiness penalty for 5 instances of inappropriate capitalization was of a similar magnitude to the penalty for 5 instances of misspelling. Note that there was a trend for the penalty of misspellings to be larger, but as a generalized rule, the 2 are similar in magnitude, and the precise difference will depend on exactly how many words and which words are capitalized or misspelled. This finding answers RQ2. Finally, with a combination of different LME models, the data show that the combination of two different types of errors had a significantly greater trustworthiness penalty than either of the error types alone, and that the effect in this study was almost perfectly additive (RQ3); thus, the effects of the combination of errors were integrative [40] rather than a simplified heuristic such as take-the-best [48], multiplicative, or affected by ceiling effects in this study. This begins to answer the question recently posed of how spelling interacts with other quality indicators on credibility [57]. To the best of our knowledge, this is the first study that was specifically designed to test and quantify these kinds of specific additive effects on trustworthiness independently of the lay participants' insight.

We also showed that the stimulus excerpts that we designed are appropriate for studies that test the trustworthiness of online health information independently of lay insight. Although our study did not preclude lay insight (ie, participants might notice that some words were misspelled), the study was not dependent on such insight, which is useful for interrogating intuitive evaluations of information (ie, cost-benefit analyses [6]). In this study, the semantic content of the excerpts engendered consistent effects on trustworthiness ratings (at least among this type of online psychology experiment cohort, see Figure 3). In addition, this is the first time that numerical values have been gathered for the isolated effects of inappropriate capitalization.

The Results in Context

This study reaffirms an earlier observation that incivility decreases message credibility [38,39]. As suggested previously, inappropriate capitalization (shouting) is histrionic and induces a strong effect of incivility on readers' subjective ratings [37]. In our study, the effect of shouting text showed a trend for eliciting a smaller trustworthiness penalty than the effect of similar quantity of misspelling. One could easily speculate about new experiments where we might change the quantities of literacy errors; our experiment used either 5 misspellings or 5 shouting phrases, but one could run experiments to titrate errors, for example, to determine the relative effects of 3 misspellings or 10 inappropriate capitalizations. Furthermore, the effects of text shouting may be moderated by whether the statements are controversial [38]. We deliberately chose statements about multiple sclerosis that would be unfamiliar to the general public. This lack of familiarity makes the content and context not particularly emotional or controversial, likely weakening the effect of inappropriate capitalization.

The debate between how accessibility (ie, cost and speed) versus information quality (ie, accuracy and presumed benefit) quantitatively affect the use of (and search for) information continues [13]. Categorical frameworks have long been proposed to provide a theoretical underpinning for the factors, grouping specific elements that influence perceptions of trustworthiness in a variety of ways. The most well-known 2-category grouping of factors affecting how people respond to communication is Petty and Cacioppo's ELM for persuasion [56], in which elements contributing to a central pathway (eg, argument quality) are complemented by seemingly less rational elements that contribute to a peripheral pathway (eg, website design) [20,41]. Another two-category persuasion model that has been used to explain online trustworthiness is Chaiken's dichotomy of heuristics versus systematic information [61,62]. In both the Heuristic-Systematic Model and the ELM, diminishing motivation/involvement (or diminishing user dependency [45]) is associated with a switch from focusing on the effortful systematic evaluation of information (quality) to low-effort heuristics (accessibility).

In purchasing decisions, the type of product affects how strongly grammar and mechanics errors affect credibility [30]. In particular, *experience goods* (nontechnical items such as body lotion that are used personally) are more affected by grammar and writing mechanics than *search goods* (technical items such as printers). The implication is that when objective signals about trustworthiness are absent, heuristics play a stronger role [56,61]. In this study, where laypersons made judgments about unfamiliar medical issues, we might expect to find a stronger response to misspelling and capitalization. A necessary future approach is to repeat this experiment with multiple sclerosis patients who would exhibit user dependence when evaluating the statements [45].

Limitations

Our study had several important constraints. We deliberately tested laypersons' judgments on unfamiliar ideas about multiple sclerosis and showed that literacy errors can have a strong effect. Nevertheless, this effect may be smaller in a cohort that is more dependent on knowing the information. In particular, if readers are dependent, then preexisting points of view and feeling of *homophily* [63] will influence perceived credibility, in a way that would not influence the general public with less interest in statements for and about multiple sclerosis.

It is also notable that in this study, the highest mean trustworthiness rating was 62.3 (0-100 scale), despite the statement being medically correct. Multiple factors may account for why this mean rating is not higher for trustworthiness. For example, participants in this psychology experiment saw the statements *in vacuo*, so that they could not verify the statement, check source credibility, or look for crowd consensus [8]. In real-life situations, other aspects of trustworthiness may dominate, and among some individuals, there may be ceiling and floor effects within this dataset, given the very wide standard deviations.

Conclusions

Incivility and literacy competence are proposed factors in how lay web users assess the trustworthiness of online health information. These results support Lederman et al's [8] theory of credibility assessment in online forums; the results also fit with Anderson's [40] description of integrative assessment of multiple cues. Here, we have shown that literacy competence errors have additive effects. The additive effects are strikingly precise. This implies that when people make seemingly rapid judgments about the trustworthiness of online text—no more than rough estimates—that their intuitive estimates seemingly add up quite accurately, at least at the population level. This completely vitiates binary models of judgment where a writer is judged as either competent or incompetent, with no further penalty for additional errors in writing mechanics [57]. The implication for writers is that for this level of errors (5 in 1 paragraph), multiple errors can add up.

Many other factors also contribute to trustworthiness, notably the argument quality of the content (logic), verification with other sources, reference credibility, and crowd consensus [8]. How these additional factors interplay with literacy competence will require further extensive research. A start would be to determine how persons with multiple sclerosis (with similar education levels to our current cohort) respond to these stimulus excerpts, as those readers would understand the information in a more self-relevant context.

Acknowledgments

The authors gratefully acknowledge Tom Ormerod and the School of Psychology at the University of Sussex for access to Qualtrics. The authors acknowledge Matthieu Raggett for information technology set up. The authors also acknowledge the BSMS Independent Research Project program for funding. Finally, the authors acknowledge the late Fayard Nicholas for the original idea on civility.

Authors' Contributions

HC, HW, GT, BM, AN, and CW conceived the experiments. HW, GT, CW, and JR conducted the experiments. HW, CJ, and GT analyzed the results. CJ performed the statistical analyses. HW, GT, and HC drafted the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

BM owns and JR worked for a commercial enterprise (Dalton Maag) that creates typefaces and graphical designs for commercial clients. AN acts as an ad hoc Scientific Consultant for Dalton Maag. This enterprise and those authors will not benefit directly or indirectly from data supporting incivility, misspelling, or shouting text. HW, GT, CW, CJ, and HC have no competing interests to declare.

Multimedia Appendix 1
Stimulus Paragraphs and errors list.
[DOC File , 51 KB - jmir_v22i6e15171_app1.doc]

Multimedia Appendix 2
Excerpts with original sources.
[DOC File , 88 KB - jmir_v22i6e15171_app2.doc]

Multimedia Appendix 3
claritinCapitalizedWords.
[XLS File (Microsoft Excel File), 83 KB - jmir_v22i6e15171_app3.xls]

Multimedia Appendix 4
Misspellings and Verifications.
[XLS File (Microsoft Excel File), 63 KB - jmir_v22i6e15171_app4.xls]

Multimedia Appendix 5
Supplementary Methods: CHERRIES.
[DOCX File , 28 KB - jmir_v22i6e15171_app5.docx]

References

1. Office for National Statistics UK. 2019. Internet Access – Households and Individuals, Great Britain: 2019 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2019 [accessed 2019-11-20]
2. Chen Y, Li C, Liang J, Tsai C. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
3. Hsieh RW, Chen L, Chen T, Liang J, Lin T, Chen Y, et al. The association between internet use and ambulatory care-seeking behaviors in Taiwan: a cross-sectional study. J Med Internet Res 2016 Dec 7;18(12):e319 [FREE Full text] [doi: 10.2196/jmir.5498] [Medline: 27927606]
4. The Health Union. 2016. Social Media for Health: What Socially-Active Patients Really Want URL: https://health-union.com/wp-content/uploads/2016/06/Social-Media-for-Health.pdf [accessed 2019-05-23] [WebCite Cache ID 78aKBnoMO]
5. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 7;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
6. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 2;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
7. Fogg B, Soohoo C, Danielson D, Marable L, Stanford J, Tauber E. How Do Users Evaluate the Credibility of Web Sites?: A Study With Over 2,500 Participants. In: Proceedings of the 2003 conference on Designing for user experiences. 2003 Presented at: DUX'03; June 6-7, 2003; San Francisco, CA. [doi: 10.1145/997078.997097]
8. Lederman R, Fan H, Smith S, Chang S. Who can you trust? Credibility assessment in online health forums. Health Policy Technol 2014 Mar;3(1):13-25. [doi: 10.1016/j.hlpt.2013.11.003]
9. Stanford J, Tauber E, Fogg B, Marable L. Advocacy - Consumer Reports. 2002. Experts vs Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites URL: https://advocacy.consumerreports.org/wp-content/uploads/2013/05/expert-vs-online-consumers.pdf [accessed 2019-05-23] [WebCite Cache ID 78aNK0qXf]
10. Brand-Gruwel S, Kammerer Y, van Meeuwen L, van Gog T. Source evaluation of domain experts and novices during web search. J Comput Assist Learn 2017 Jan 5;33(3):234-251. [doi: 10.1111/jcal.12162]
11. Sillence E, Briggs P, Harris P, Fishwick L. Going online for health advice: changes in usage and trust practices over the last five years. Interact Comput 2007 May;19(3):397-406 [FREE Full text] [doi: 10.1016/j.intcom.2006.10.002]
12. England CY, Nicholls AM. Advice available on the internet for people with coeliac disease: an evaluation of the quality of websites. J Hum Nutr Diet 2004 Dec;17(6):547-559. [doi: 10.1111/j.1365-277X.2004.00561.x] [Medline: 15546433]
13. Woudstra L, van den Hooff B, Schouten A. The quality versus accessibility debate revisited: a contingency perspective on human information source selection. J Assn Inf Sci Tec 2015 May 13;67(9):2060-2071. [doi: 10.1002/asi.23536]
14. Zimmermann M, Jucks R. How experts' use of medical technical jargon in different types of online health forums affects perceived information credibility: randomized experiment with laypersons. J Med Internet Res 2018 Jan 23;20(1):e30 [FREE Full text] [doi: 10.2196/jmir.8346] [Medline: 29362212]
15. Kata A. Anti-vaccine activists, web 2.0, and the postmodern paradigm-an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
16. Borzekowski DL, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: a content analysis of pro-eating disorder web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]
17. Flynn D, Nyhan B, Reifler J. The nature and origins of misperceptions: understanding false and unsupported beliefs about politics. Adv Political Psychol 2017 Jan 26;38:127-150. [doi: 10.1111/pops.12394]
18. Pickard V. The Political Economy of Communication. 2017. Media Failures in the Age of Trump URL: http://www.polecom.org/index.php/polecom/article/view/74/264 [accessed 2018-08-26]
19. Hobbs J, Kittler A, Fox S, Middleton B, Bates DW. Communicating health information to an alarmed public facing a threat such as a bioterrorist attack. J Health Commun 2004;9(1):67-75. [doi: 10.1080/10810730490271638] [Medline: 14761834]
20. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
21. Albuja AF, Sanchez DT, Gaither SE. Fluid racial presentation: perceptions of contextual 'passing' among biracial people. J Exp Soc Psychol 2018 Jul;77:132-142. [doi: 10.1016/j.jesp.2018.04.010]
22. Rieh SY, Danielson DR. Credibility: a multidisciplinary framework. Ann Rev Info Sci Tech 2008 Oct 24;41(1):307-364. [doi: 10.1002/aris.2007.1440410114]

23. Gray R, Fitch M, Davis C, Phillips C. Breast cancer and prostate cancer self‐help groups: reflections on differences. Psycho‐Oncology 1996;5(2):137-142. [doi: 10.1002/(sici)1099-1611(199606)5:2<137::aid-pon222>3.3.co;2-5]
24. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc Sci Med 2006 May;62(10):2577-2590. [doi: 10.1016/j.socscimed.2005.11.016] [Medline: 16361029]
25. BBC News Magazine. 2012. The Cruellest of Internet Hoaxes URL: https://www.bbc.co.uk/news/magazine-18282277 [accessed 2019-05-23] [WebCite Cache ID 78aSUcf7v]
26. Metzger MJ. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. J Am Soc Inf Sci 2007 Nov;58(13):2078-2091. [doi: 10.1002/asi.20672]
27. Kornfield R, Smith KC, Szczypka G, Vera L, Emery S. Earned media and public engagement with CDC's 'tips from former smokers' campaign: an analysis of online news and blog coverage. J Med Internet Res 2015 Jan 20;17(1):e12 [FREE Full text] [doi: 10.2196/jmir.3645] [Medline: 25604520]
28. Gupta S, MacLean DL, Heer J, Manning CD. Induced lexico-syntactic patterns improve information extraction from online medical forums. J Am Med Inform Assoc 2014;21(5):902-909 [FREE Full text] [doi: 10.1136/amiajnl-2014-002669] [Medline: 24970840]
29. Foufi V, Gaudet-Blavignac C, Chevrier R, Lovis C. De-identification of medical narrative data. Stud Health Technol Inform 2017;244:23-27. [Medline: 29039370]
30. Ketron S. Investigating the effect of quality of grammar and mechanics (QGAM) in online reviews: the mediating role of reviewer crediblity. J Bus Res 2017 Dec;81:51-59. [doi: 10.1016/j.jbusres.2017.08.008]
31. Schindler RM, Bickart B. Perceived helpfulness of online consumer reviews: the role of message content and style. J Consumer Behav 2012 Mar 21;11(3):234-243. [doi: 10.1002/cb.1372]
32. Figueredo L, Varnhagen CK. Didn't you run the spell checker? Effects of type of spelling error and use of a spell checker on perceptions of the author. Read Psychol 2005 Sep;26(4-5):441-458. [doi: 10.1080/02702710500400495]
33. Hargittai E. Hurdles to information seeking: spelling and typographical mistakes during users' online behavior. J Assoc Inf Syst 2006 Jan;7(1):52-67. [doi: 10.17705/1jais.00076]
34. Morin-Lessard E, McKelvie SJ. Does writeing rite matter? Effects of textual errors on personality trait attributions. Curr Psychol 2017 Mar 24;38(1):21-32. [doi: 10.1007/s12144-017-9582-z]
35. Vignovic JA, Thompson LF. Computer-mediated cross-cultural collaboration: attributing communication errors to the person versus the situation. J Appl Psychol 2010 Mar;95(2):265-276. [doi: 10.1037/a0018628] [Medline: 20230068]
36. Sobieraj S, Berry JM. From incivility to outrage: political discourse in blogs, talk radio, and cable news. Political Commun 2011 Feb 9;28(1):19-41. [doi: 10.1080/10584609.2010.542360]
37. Gervais BT. Incivility online: affective and behavioral reactions to uncivil political posts in a web-based experiment. J Inf Technol Politics 2015 Jan 14;12(2):167-185. [doi: 10.1080/19331681.2014.997416]
38. Graf J, Erba J, Harn R. The role of civility and anonymity on perceptions of online comments. Mass Commun Soc 2017 Feb 21;20(4):526-549. [doi: 10.1080/15205436.2016.1274763]
39. Thorson K, Vraga E, Ekdale B. Credibility in context: how uncivil online commentary affects news credibility. Mass Commun Soc 2010 Jun 3;13(3):289-313. [doi: 10.1080/15205430903225571]
40. Anderson N. Unified psychology based on three laws of information integration. Rev Gen Psychol 2013 Jun;17(2):125-132 [FREE Full text] [doi: 10.1037/a0032921]
41. Fogg B. Prominence-Interpretation Theory: Explaining How People Assess Credibility Online. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems. 2003 Presented at: CHI EA'03; April 5-10, 2003; Florida, USA p. 722-723. [doi: 10.1145/765891.765951]
42. Wanas N, El-Saban M, Ashour H, Ammar W. Automatic Scoring of Online Discussion Posts. In: Proceedings of the 2nd ACM Workshop on Information Credibility on the Web. 2008 Presented at: WICOW'08; October 30, 2008; Napa Valley, California, USA p. 19-26. [doi: 10.1145/1458527.1458534]
43. Castillo C, Mendoza M, Poblete B. Predicting information credibility in time-sensitive social media. Internet Res 2013 Oct 14;23(5):560-588. [doi: 10.1108/IntR-05-2012-0095]
44. Weerkamp W, de Rijke M. Credibility-inspired ranking for blog post retrieval. Inf Retrieval 2012 Feb 11;15(3-4):243-277. [doi: 10.1007/s10791-011-9182-8]
45. Johnson F, Rowley J, Sbaffi L. Modelling trust formation in health information contexts. J Inf Sci 2015 Mar 25;41(4):415-429. [doi: 10.1177/0165551515577914]
46. Pachur T, Bröder A. Judgment: a cognitive processing perspective. Wiley Interdiscip Rev Cogn Sci 2013 Nov;4(6):665-681. [doi: 10.1002/wcs.1259] [Medline: 26304271]
47. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974 Sep 27;185(4157):1124-1131. [doi: 10.1126/science.185.4157.1124] [Medline: 17835457]
48. Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev 1996 Oct;103(4):650-669. [doi: 10.1037/0033-295x.103.4.650] [Medline: 8888650]
49. Gigerenzer G, Goldstein DG. The recognition heuristic: a decade of research. Judgm Decis Mak 2011;6:100-121 [FREE Full text] [doi: 10.1093/acprof:oso/9780199390076.003.0008]

50. Bennis W, Katsikopoulos K, Goldstein D, Dieckmann AN, Berg N. Designed to fit minds. In: Todd PM, Gigerenzer G, editors. Ecological Rationality: Intelligence in the World. New York, USA: Oxford University Press; 2012:409-427.
51. Oreg S, Bayazit M. Prone to bias: development of a bias taxonomy from an individual differences perspective. Rev Gen Psychol 2009 Sep;13(3):175-193. [doi: 10.1037/a0015656]
52. Koriat A. Metacognition: decision-making processes in self-monitoring and self-regulation. In: Keren G, Wu G, editors. The Wiley Blackwell Handbook of Judgment and Decision Making. Chichester, UK: John Wiley & Sons; 2015:356-379.
53. Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc Natl Acad Sci U S A 2019 Sep 24;116(39):19251-19257 [FREE Full text] [doi: 10.1073/pnas.1821936116] [Medline: 31484770]
54. Orne MT. Demand characteristics and the concept of quasi-controls. In: Rosenthal R, Rosnow RL, editors. Artifacts in Behavioral Research. Oxford, UK: Oxford University Press; 2009:110-137.
55. Krefting L. Rigor in qualitative research: the assessment of trustworthiness. Am J Occup Ther 1991 Mar;45(3):214-222. [doi: 10.5014/ajot.45.3.214] [Medline: 2031523]
56. Petty RE, Cacioppo JT. The effects of involvement on responses to argument quantity and quality: central and peripheral routes to persuasion. J Pers Soc Psychol 1984;46(1):69-81. [doi: 10.1037/0022-3514.46.1.69]
57. Sauls M. TigerPrints: Clemson University Research. 2018. Perceived Credibility of Information on Internet Health Forums URL: https://tigerprints.clemson.edu/all_dissertations/2110 [accessed 2019-05-24]
58. Oleson D. World Bank Open Data. 2013. Discovering Drug Side Effects With Crowdsourcing URL: https://data.world/crowdflower/claritin-twitter/workspace/file?filename=1384367161_claritin_october_twitter_side_effects-1.csv [accessed 2019-11-23]
59. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
60. Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics 2000 Jun;56(2):645-646. [doi: 10.1111/j.0006-341x.2000.00645.x] [Medline: 10877330]
61. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]
62. Sundar S. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ, Flanagin AJ, editors. Digital Media, Youth, and Credibility. Cambridge, MA: The MIT Press; 2008:73-100.
63. Wang Z, Walther JB, Pingree S, Hawkins RP. Health information, credibility, homophily, and influence via the internet: web sites versus discussion groups. Health Commun 2008 Jul;23(4):358-368. [doi: 10.1080/10410230802229738] [Medline: 18702000]

Abbreviations

ELM: Elaboration Likelihood Model
LME: linear mixed effects
RQ: research question

Edited by G Eysenbach; submitted 26.06.19; peer-reviewed by M Tkáč, L Sbaffi, M Reynolds; comments to author 19.07.19; revised version received 06.12.19; accepted 15.12.19; published 10.06.20.

Please cite as:
Witchel HJ, Thompson GA, Jones CI, Westling CEI, Romero J, Nicotra A, Maag B, Critchley HD
Spelling Errors and Shouting Capitalization Lead to Additive Penalties to Trustworthiness of Online Health Information: Randomized Experiment With Laypersons
J Med Internet Res 2020;22(6):e15171
URL: https://www.jmir.org/2020/6/e15171
doi:10.2196/15171
PMID:32519676

Original Paper

Web-Based Health Information Following the Renewal of the Cervical Screening Program in Australia: Evaluation of Readability, Understandability, and Credibility

Olivia A Mac[1], BSc, MPH; Amy Thayre[1], BSc, MPH; Shumei Tan[1], BSc, MPH; Rachael H Dodd[1], BSc, MPH, PhD

[1]School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Corresponding Author:
Rachael H Dodd, BSc, MPH, PhD
School of Public Health
Faculty of Medicine and Health
The University of Sydney
Edward Ford Building
Sydney, 2006
Australia
Phone: 61 293515102
Email: rachael.dodd@sydney.edu.au

Abstract

Background: Three main changes were implemented in the Australian National Cervical Screening Program (NCSP) in December 2017: an increase in the recommended age to start screening, extended screening intervals, and change from the Papanicolaou (Pap) test to primary human papillomavirus screening (cervical screening test). The internet is a readily accessible source of information to explain the reasons for these changes to the public. It is important that web-based health information about changes to national screening programs is accessible and understandable for the general population.

Objective: This study aimed to evaluate Australian web-based resources that provide information about the changes to the cervical screening program.

Methods: The term *cervical screening* was searched in 3 search engines. The first 10 relevant results across the first 3 pages of each search engine were selected. Overall, 2 authors independently evaluated each website for readability (Flesch Reading Ease [FRE], Flesch-Kincaid Grade Level, and *Simple Measure of Gobbledygook* [SMOG] index), quality of information (Patient Education Materials Assessment Tool [PEMAT] for printable materials), credibility (*Journal of the American Medical Association* [JAMA] benchmark criteria and presence of Health on the Net Foundation code of conduct [HONcode] certification), website design, and usability with 5 simulation questions to assess the relevance of information. A descriptive analysis was conducted for the readability measures, PEMAT, and the JAMA benchmark criteria.

Results: Of the 49 websites identified in the search, 15 were eligible for inclusion. The consumer-focused websites were classed as *fairly difficult to read* (mean FRE score 51.8, SD 13.3). The highest FRE score (easiest to read) was 70.4 (*Cancer Council Australia Cervical Screening Consumer Site*), and the lowest FRE score (most difficult to read) was 33.0 (*NCSP Clinical Guidelines*). A total of 9 consumer-focused websites and 4 health care provider–focused websites met the recommended threshold (sixth to eighth grade; SMOG index) for readability. The mean PEMAT understandability scores were 87.7% (SD 6.0%) for consumer-focused websites and 64.9% (SD 13.8%) for health care provider–focused websites. The mean actionability scores were 58.1% (SD 19.1%) for consumer-focused websites and 36.7% (SD 11.0%) for health care provider–focused websites. Moreover, 9 consumer-focused and 3 health care provider–focused websites scored above 70% for understandability, and 2 consumer-focused websites had an actionability score above 70%. A total of 3 websites met all 4 of the JAMA benchmark criteria, and 2 websites displayed the HONcode.

Conclusions: It is important for women to have access to information that is at an appropriate reading level to better understand the implications of the changes to the cervical screening program. These findings can help health care providers direct their patients toward websites that provide information on cervical screening that is written at accessible reading levels and has high understandability.

(J Med Internet Res 2020;22(6):e16701) doi:10.2196/16701

KEYWORDS

cervical screening; internet; consumer health information; Australia; papillomavirus infections

Introduction

Background

In Australia, the latest figures for cervical screening from 2017 to 2018 show a 2-year participation rate of 53% for women aged 25 to 69 years [1] and an overall incidence of cervical cancer at 7 cases per 100,000 women [2]. Prevention of cervical cancer through the Australian National Cervical Screening Program (NCSP) was first introduced in 1991 and screened women aged 18 to 69 years every 2 years using cytology-based screening (Pap smear). The program was renewed in December 2017, and women aged 25 to 74 years are screened every 5 years using primary human papillomavirus (HPV) screening (Cervical Screening Test) [3]. This renewal was based on a greater understanding of the natural history of HPV and cervical cancer, successful uptake of the HPV vaccination leading to a subsequent reduction in vaccine-related HPV types, evidence that the HPV test is more sensitive than the Pap smear, and economic modeling demonstrating HPV screening to be more cost-effective [3]. Recent modeling studies have predicted that the new program will reduce the incidence and mortality of cervical cancer in vaccinated women by 31% and 36%, respectively [3,4]. A limited understanding of the rationale behind these changes has been demonstrated by the general population [5,6]. This highlights a need for information explaining these changes that is easy to understand and access to not undermine the confidence women have in the screening program.

Almost 80% of Australians now use the internet as a source of health information, suggesting that the internet could be a powerful tool to educate and inform readers, with the potential to alleviate anxiety or concern [7]. As health models lean toward greater patient empowerment, patients may feel a greater sense of responsibility for their health care. However, because of the largely unrestricted nature of the internet and limited governance, there is a risk that users may be exposed to inaccurate, unreliable, biased, or potentially harmful information [8]. Exposure to information that is not presented in an accessible manner may cause unnecessary anxiety and distress. This could result in inappropriate care or people ignoring evidence-based recommendations because of being exposed to contradictory information [9].

Moreover, health information is only useful if it can be understood by the target population. The ability to read and understand written text is key for the comprehension of health information. Readability tools can measure the reading ability needed to understand the information presented. South Australia Health recommends that the readability level for health information is grade 8 (12-14 years old with 8 years of Australian education) [10,11]. In the United States, the National Institutes of Health and the American Medical Association recommend patient education materials to be written at or below the sixth-grade reading level (age 11 or 12 years), whereas the Joint Commission recommends a fifth-grade level (age 10 or 11 years) or lower [12]. There are limitations to using only readability measures to evaluate health care information [13], so it is important to ensure that web-based health information is of sufficient quality and is suitable to people from diverse backgrounds and varying levels of health literacy. Web-based health information may also require higher literacy levels than printed patient education materials, as previous studies have demonstrated that web-based patient education materials are often written above the recommended reading level of grades 6 to 8 [14-17], and these may be perceived as more difficult than print materials [13].

There is little to no value of health information that is trustworthy and credible if it cannot be easily understood and acted upon by the general population. A commonly cited disadvantage of web-based health information is the inability of consumers to evaluate the quality of websites [9,18]. In addition to being able to read health care information, consumers need to be able to understand the information presented, evaluate the credibility of websites, and find the information they are looking for.

Objectives

To the best of our knowledge, no previous research has evaluated the readability of health information provided on websites about cervical screening in Australia, particularly regarding the changes to the NCSP. Therefore, this study aimed to evaluate the information available on the internet regarding the renewal of the NCSP in Australia, with a particular focus on website readability, understandability, design, credibility, and usability.

Methods

Identification of Websites

In April 2019, the term *cervical screening* was searched for in Australian versions of the 3 most popular search engines: *Google, Yahoo, and Bing*. The incognito window on Google Chrome was used to conduct the search, and browser history, cache, and cookies were cleared before running the search to ensure that previous searches would not impact the search results. The first 10 relevant results from the first 3 pages of results from each search engine were selected. Relevant websites (in the English language and related to cervical screening) were identified from each search engine, and any duplicate websites were removed and noted. Websites were excluded if they were advertisements, news reports, Wikipedia pages, social media pages, online discussion forums, blogs, videos, books, articles, or private websites.

Measures

Readability

Readability was measured using Flesch Reading Ease (FRE), Flesch-Kincaid Grade Level (FKGL) [19], and the Simple Measure of Gobbledygook (SMOG) index [20].

Flesch Reading Ease and Flesch-Kincaid Grade Level

FRE uses a formula based on the average number of syllables, words per sentence, and the number of sentences to generate a readability score between 0 and 100 (Multimedia Appendix 1). Higher scores indicate greater ease of comprehension. A readability score above 60 is considered easy to follow by the general population [19]. The FRE has high reproducibility and correlation with other readability measures [21].

FKGL is a modified version of the FRE that generates the average US grade level required to understand the information. For example, an FKGL score of 8 indicates that the text can be understood by readers who have completed the equivalent of US grade 8 (Australian year 8; approximate age of 12-14 years) [19,22].

Simple Measure of Gobbledygook Index

The SMOG index assesses 10 consecutive sentences at the beginning, middle, and end of the relevant text and counts the number of polysyllabic words in each sentence [20]. These results are entered into a formula to establish the required grade level. A SMOG score of 3 to 8, 9 to 12, and 13 to 18 indicates that completion of primary, secondary, or tertiary education, respectively, is required to understand the information. For both FKGL and the SMOG index, a higher score indicates that a higher education level is required to understand the information.

The combination of our chosen readability measures (FRE, FKGL, and the SMOG index) is considered optimal, as they have been validated in the context of web-based health information and have high reliability for analyzing biomedical information [21].

Understandability and Actionability

The websites were evaluated for understandability and actionability using the validated Patient Education Materials Assessment Tool (PEMAT) for printable materials [23]. The PEMAT consists of 2 subscales: (1) understandability, which is a measure of the extent to which patient education materials can be understood by people of varying health literacy levels and diverse backgrounds; and (2) actionability, which measures how well a health consumer is able to identify what action to take based on the information provided [23]. Items are given a score of either 0 (disagree) or 1 (agree), with some items having a *not applicable* option. Final scores are calculated as a percentage of *agree* responses for all items, excluding those rated as *not applicable*. Higher percentages indicate higher understandability or actionability. A score higher than 70% indicates that materials are understandable and/or actionable [18]. As actionability was less relevant to this context, it was given less weight when determining the best overall website.

Website Credibility

Health on the Net Foundation Code of Conduct

The Health on the Net (HON) Foundation created a code of conduct that allows certified websites to display the Health on the Net Foundation code of conduct (HONcode) logo as a seal of approval [24]. The HONcode is recognized as an ethical code for websites and is based on 8 principles: authority, complementarity, privacy policy, attribution and date, justifiability, transparency, financial disclosure, and advertising policy. Website developers or information providers can apply for membership and request HONcode certification free of charge. The HONcode seal of approval is given to websites that comply with the principles mentioned earlier. Each website was evaluated for the presence of the HONcode logo.

Journal of the American Medical Association Benchmark

The *Journal of the American Medical Association* (JAMA) benchmark criteria enable the reader to easily discredit websites that lack reliability and transparency [25]. The criteria are as follows: (1) authorship—provides details about authors, contributors, affiliations, and credentials; (2) attribution—all references, sources, and copyright information to be provided; (3) currency—provision of the dates that content was updated; and (4) disclosure—website ownership, sponsorship, advertising policies, and potential conflicts of interest are prominently disclosed [25]. Websites were evaluated against each of the 4 benchmarks and given 1 point for each criterion met (final score 0-4).

Website Design

We used information on readability from the National Institute of Adult Continuing Education (NIACE) [26] to evaluate the design features that make the content of health care information more accessible and easier to understand (Multimedia Appendix 2). We analyzed the use of clear and distinct font styles (NIACE recommends fonts similar to Helvetica) and the adaptability of text (the ability to change the text size) and responsive web design (ie, the website displays differently according to the type of device, screen size, and orientation). Google Chrome features a *toggle device toolbar*, which allows the user to view how a website will be displayed on different screen sizes. We tested the responsiveness of each website by emulating different device types. We analyzed whether the font size appropriately reflected the purpose of the text (eg, larger text size for titles and headings). We also assessed whether the information was broken into *chunks* of text separated by white space as well as the presence and relevance of illustrations.

Usability: Simulation Questions

To emulate real life, 2 researchers, separately and independently, attempted to answer the following questions that women are likely to have related to the renewed NCSP: (1) Why did they change the interval from 2 to 5 years? (2) Why did they change the test? (3) Why did they change the age? (4) What are the benefits of the new test? and (5) Who should have the new test? Each researcher assessed whether the question was answered, and if yes, how well it was answered and how long it took to find the answer. The average time taken by both researchers was reported.

Analysis

Readability statistics were obtained by copying and pasting each website URL into a web-based analysis tool, the *Readability Test Tool* [27], which uses an algorithm to calculate readability measures. If there were several pages on a website, statistics for each page were obtained, and the mean was calculated. An exception to this was if a website had a *publications and resources* page; the statistics for this page were not calculated. The analysis was stratified by the intended audience: consumer focused *versus* health care provider focused. Websites were categorized according to their primary target audience (ie, websites with most content aimed toward consumers were treated as consumer focused).

The quality of websites was evaluated by 2 independent researchers using Microsoft Excel. A third researcher independently coded 40% (6/15) of the websites. Descriptive statistics were calculated using IBM SPSS version 25. Means and SDs were calculated for understandability and actionability (PEMAT), readability statistics (FRE, FKGL, and SMOG), and JAMA benchmark criteria. Interrater reliability between the 2 reviewers was assessed using the Cohen kappa coefficient. Correlations between PEMAT and readability measures were calculated using the Pearson correlation coefficient. The Kruskal-Wallis test was conducted to detect differences in mean readability, understandability, actionability, and JAMA benchmark scores between consumer-focused and health care provider–focused websites.

Ethical Approval

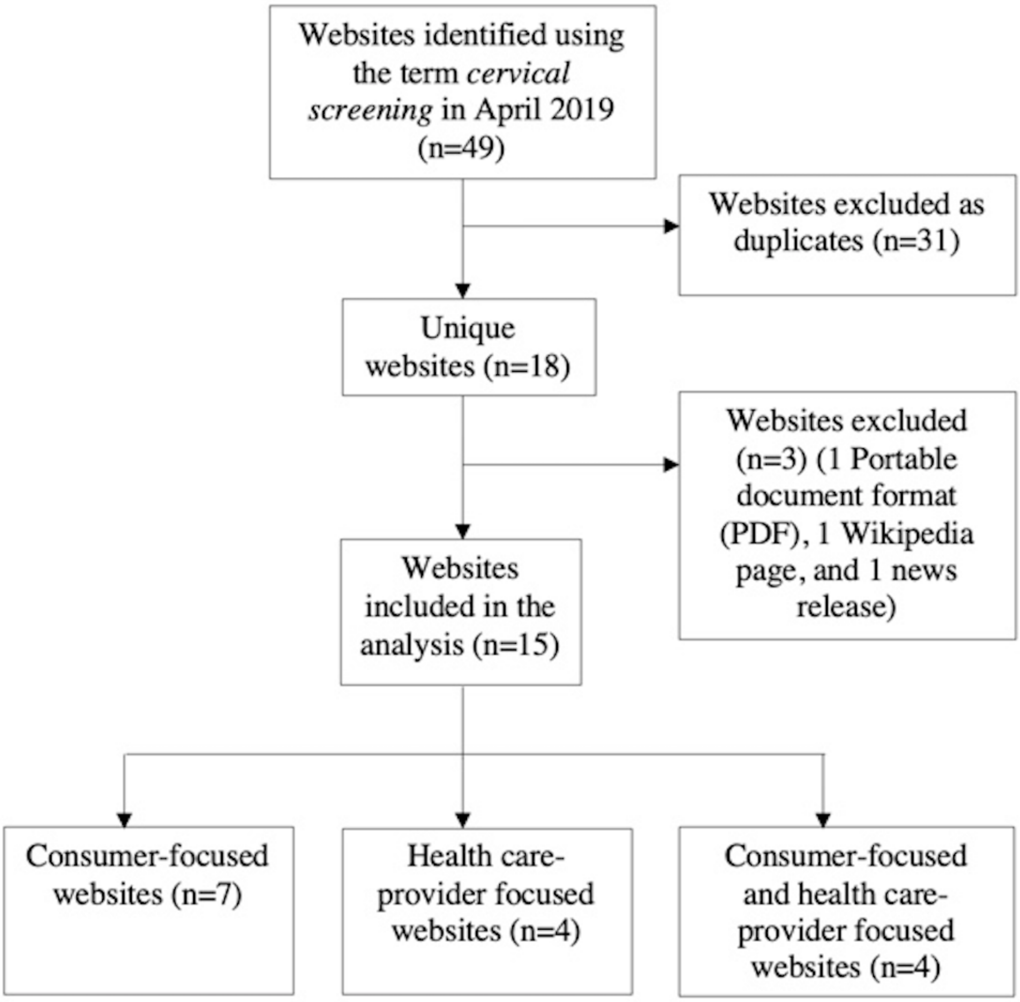
Ethical approval was not required, as the websites were in the public domain, and no human participants were involved.

Results

Search Results

A total of 49 websites were identified using *Google* (n=16), *Yahoo* (n=15), *and Bing* (n=18; Multimedia Appendix 3). After applying the exclusion criteria, 15 unique websites remained and were included in the evaluation (Figure 1). Duplicates (n=31) were excluded; therefore, a greater number of eligible websites were included from *Google*: *Google* (n=10), *Yahoo* (n=1), and *Bing* (n=4; full URLs can be found in Multimedia Appendix 3). The included websites were federal (*Cancer Australia)* or state government–owned (eg, *Cancer Institute NSW*), government-funded (*Health Direct)* or nongovernmental organizations (*Cancer Council).* Of the 15 included websites, 10 were targeted toward consumers, and 5 were targeted toward health care providers, including *National Prescribing Service (NPS) MedicineWise* that consisted of training modules for health care providers. In addition, 4 websites had information for both consumers and health care providers but were categorized according to the primary target audience. Abbreviations, target audience, and organizations of websites included in the analysis are described in Multimedia Appendix 4.

Figure 1. Flow diagram of inclusion and exclusion of websites for evaluation.



Readability

The mean FRE of consumer-focused websites was 51.8 (SD 13.3; Table 1), which is considered *fairly difficult to read*. Health care provider–focused websites were considered *difficult to read*, with a mean FRE of 43.7 (SD 7.3). Of the consumer-focused websites, the *Cancer Council Australia Cervical Screening Consumer Site* had the highest FRE (70.4; Table 2), and the *Cancer Council Australia Main Site* had the lowest FRE (37.8). Consumer-focused websites had a mean reading grade level of 7.8 (SD 1.5), compared with a mean reading grade level of 8.1 (SD 1.4; Table 1) for health care provider–focused sites. *Health Direct* was the only website to score below the US-recommended sixth-grade level using the FKGL (Table 2). The mean SMOG index of consumer-focused sites was 6.7 (SD 1.2), compared with a mean SMOG index of 7.5 (SD 0.7; Table 1) for health care provider–focused sites. The SMOG index ranged from 4.6 (*Cancer Australia Cervical Screening)* to 9.1 *(Victoria Cervical Screening Program [CSP]).* Overall, 13 out of 15 (87%) of all included websites (9 consumer focused and health care provider–focused) scored between the sixth and eighth grade reading level using the SMOG index. *Victoria CSP* had the highest grade level (FKGL and SMOG) and the lowest reading ease (FRE). The differences in the FRE, FKGL, and SMOG index scores of consumer-focused and health care provider–focused websites were not statistically significant.

Table 1. Descriptive statistics of included websites by target audience.

Target audience of website	Flesch Reading Ease[a] (0-100)	Flesch-Kincaid Grade Level[b] (0-12)	SMOG[c] index (3-18)	PEMAT[d] understandability[e] (0-100)	PEMAT actionability[e] (0-100)	JAMA[f] benchmark criteria[g] (0-4)
Consumers (n=10)						
Mean (SD)	51.8 (13.3)	7.8 (1.5)	6.7 (1.2)	87.7 (6.0)	58.1 (19.1)	2.4 (0.8)
Median	50.0	7.6	6.8	90.0	60.0	2.0
Range	36.4-70.4	5.9-10.4	4.6-9.1	75.0-93.0	20.0-93.3	1.0-4.0
Health care providers (n=5)						
Mean (SD)	43.7 (7.3)	8.8 (1.1)	7.5 (0.7)	64.9 (13.8)	36.7 (11.0)	2.6 (1.3)
Median	44.4	8.6	7.3	69.2	40.0	2.0
Range	33.0-53.3	7.8-10.7	6.7-8.7	64.3-77.8	20.0-40.0	2.0-4.0
All (n=15)						
Mean (SD)	48.9 (11.9)	8.1 (1.4)	7.0 (1.1)	79.0 (14.5)	51.0 (18.2)	2.3 (1.1)
Median	45.0	8.2	6.9	83.6	85.0	2.0
Range	33.0-70.4	5.9-10.7	4.6-9.1	64.3-93.0	20.0-60.0	1.0-4.0

[a] A Flesch reading ease score of 60 or higher is considered easy to read by the general public.

[b] Lower Flesch-Kincaid grade level and SMOG index indicate that content is easier to read. It is recommended that information is written between grades 6 and 8 (with grade 6 being easier to read).

[c] SMOG: Simple Measure of Gobbledygook.

[d] PEMAT: Patient Education Materials Assessment tool.

[e] A PEMAT score of 70 or higher indicates that content is understandable and/or actionable.

[f] JAMA: Journal of the American Medical Association.

[g] Scores represent the mean number of JAMA benchmark criteria satisfied. Websites were evaluated against each of the 4 benchmarks (authorship, attribution, currency, and disclosure) and given 1 point for each criterion met.

Table 2. Readability, Patient Education Materials Assessment Tool for Printable Materials (PEMAT-P) scores, and credibility of included websites.

Search engine and website	Readability			PEMAT[a]		Credibility	
	Flesch Reading Ease[b] (0-100)	Flesch-Kincaid Grade Level[c] (0-12)	SMOG[d] index[c] (3-18)	Understandability[e] (0-100)	Actionability[e] (0-100)	HONcode[f] presence (yes/no)	JAMA[g] benchmark criteria (Au, At, D, and C)[h]
Google							
Consumer focused							
NCSP[i]	60.6	6.9	6.9	93.0	83.3	No	D, C
CCA[j] main site	37.8	8.9	5.7	81.8	20.0	No	Au, At, D, C
CCA cervical screening consumer site	70.4	6.3	6.8	85.7	60.0	No	Au, At, D
Health Direct	68.1	5.9	6.6	92.3	40.0	Yes	At, D, C
Jean Hailes	56.2	6.9	6.1	87.5	80.0	No	D, C
WA[k] CSP[l]	60.4	6.8	6.9	91.7	60.0	No	D, C
NSW[m] CSP	49.5	7.6	6.4	92.3	60.0	No	At
Victoria CSP	36.4	10.4	9.1	90.0	60.0	No	At, C
Queensland CSP (A)	44.4	8.2	6.7	69.2	40.0	No	D, C
Health care provider focused							
RACGP[n]	42.2	8.8	7.2	64.3	50.0	No	Au, At, D, C
Yahoo							
Consumer focused							
Cancer Australia cervical screening	40.7	8.1	4.6	75.0	60.0	No	Au, C
Bing							
Health care provider focused							
Queensland CSP (B)	39.2	8.8	6.5	82.4	60.0	No	D, C
NPS[o] MedicineWise	53.3	7.8	7.5	77.8	20.0	Yes	D, C
NCSP clinical guidelines	33.0	10.7	8.7	71.4	33.3	No	Au, At, D, C
NCSR[p]	45.5	8.6	7.3	41.7	40.0	No	Au

[a] PEMAT: Patient Education Materials Assessment Tool.

[b] A Flesch Reading Ease score of 60 or higher is considered easy to read by the general public.

[c] Lower Flesch-Kincaid Grade Level and SMOG index indicate content is easier to read. It is recommended that information is written between grade 6 and 8 (with grade 6 being easier to read).

[d] SMOG: Simple Measure of Gobbledygook.

[e] A PEMAT score of 70 or higher indicates content is understandable and/or actionable.

[f] HONCode: Health on the Net Foundation code of conduct.

[g] JAMA: Journal of the American Medical Association.

[h] Websites were evaluated against each of the 4 JAMA benchmark criteria: authorship (Au), attribution (At), disclosure (D), and currency (C).

[i] NCSP: National Cervical Screening Program.

[j] CCA: Cancer Council Australia.

[k] WA: Western Australia.

[l] CSP: cervical screening program.

[m] NSW: New South Wales.

[n] RACGP: Royal Australian College of General Practitioners.

[o] NPS: National Prescribing Service.

[p] NCSR: National Cancer Screening Register.

Understandability and Actionability

The mean understandability score of consumer-focused websites was significantly higher than that of health care provider–focused websites (87.7%, SD 6.0% vs 64.9%, SD 13.8%; X^2_1=6.9; *P*=.01). Overall, 90% (9/10) of consumer-focused websites and 80% (4/5) of health care-provider focused websites met or exceeded the 70% threshold for understandability. The mean PEMAT actionability score was significantly higher in consumer-focused websites than in health care provider–focused websites (58.1%, SD 19.1% vs 36.7%, SD 11.0%; X^2_1=4.8; *P*=.04). Two consumer-focused websites met the threshold for actionability, whereas no health care provider-focused website was found to have actionable information. The *NCSP Main Site* and *Jean Hailes* were the only 2 websites to score over 70% in both the understandability and actionability domains. Interrater reliability was substantial for PEMAT ratings with 0.73 for understandability and 0.75 for actionability.

All 15 websites made their purpose evident and defined terms; however, only 13% (2/15; Table 3) of websites were judged as using visual aids whenever possible, and 53% (8/15) of the websites were judged to use the active voice. All consumer-focused websites, and 60% (3/5; Table 3) of health care provider-focused websites identified at least one action for the reader to take. Overall, 53% (8/15) of websites broke down the action into explicit steps, and 7% (1/15) of websites provided tangible tools whenever possible to help the reader take action. A correlation analysis was conducted between the FRE and PEMAT understandability to determine if websites that were easier to read were also easier to understand. There was no significant correlation found between reading ease and understandability.

Table 3. Percentage of *agree* responses on Patient Education Materials Assessment Tool (PEMAT) for Printable Materials items of the 15 included websites.

PEMAT[a] items	Agree, n (%)[b]		
	Consumer-focused websites	Health care provider–focused websites	All websites
Understandability			
Content			
Makes its purpose completely evident	10 (100)	5 (100)	15 (100)
No distracting information	10 (100)	5 (100)	15 (100)
Word choice and style			
Common everyday language	10 (100)	3 (60)	13 (87)
Medical terms are defined and used only to familiarize readers	10 (100)	3 (60)	13 (87)
Active voice	6 (60)	2 (40)	8 (53)
Use of numbers			
Numbers are clear and easy to understand[c]	5(100)	3(75)	8 (73)
Does not expect readers to do calculations	10 (100)	5 (100)	10 (100)
Organization			
Chunks information into short sections[d]	9 (100)	3 (100)	12 (100)
Sections have informative headings[d]	9 (100)	3 (100)	12 (100)
Presents information in a logical sequence	10 (100)	5(100)	15 (100)
Provides a summary[d]	6 (67)	2 (67)	8 (67)
Layout and design			
Provides visual cues whenever possible	9 (90)	4 (80)	13 (87)
Use of VA[e]			
Uses VA whenever possible	2 (20)	0 (0)	2 (13)
VA reinforce rather than distract[f]	4 (100)	1 (100)	5 (100)
VA have clear titles and captions[f]	2 (50)	1 (100)	3 (60)
VA are clear and uncluttered[f]	4 (100)	1 (100)	5 (100)
Tables are simple with short, clear row and column headings[g]	3 (100)	1 (100)	4 (100)
Actionability			
Identifies at least one action for the user	10 (100)	3 (60)	13 (87)
Addresses the user directly	9 (90)	3 (60)	11 (73)
Breaks down actions into explicit steps	6 (60)	2 (40)	8 (53)
Provides tangible tools whenever it could help	0 (0)	1 (20)	1 (7)
Instructions and examples for calculations[h]	N/A[i]	N/A	N/A
Explains how to use the charts, diagrams etc[j]	2 (100)	0 (0)	2 (67)
Uses VA whenever possible to help take action	2 (20)	2 (40)	4 (27)

[a] PEMAT: patient education materials assessment tool.

[b] Agree (%) was calculated by the following formula: total number of *agrees*/total number of applicable websites.

[c] Not applicable for 4 websites.

[d] Not applicable for 3 websites.

[e] VA: visual aids.

[f] Not applicable for 10 websites.

[g] Not applicable for 11 websites.

[h] Not applicable for all 15 websites.

[i] N/A: not applicable.

[j] Not applicable for 12 websites.

Credibility: Journal of the American Medical Association Benchmark Criteria and Health on the Net Foundation Code of Conduct

Overall, 3 websites met all 4 of the JAMA benchmark criteria (*Cancer Council Australia Main Site*, *NCSP Clinical Guidelines*, and *the Royal Australian College of General Practitioners [RACGP]).* The mean number of JAMA benchmark criteria satisfied was 2.4, ranging from 1 to 4. Overall, currency was the most well-adhered principle, with 8 of 10 consumer-focused and 4 of 5 health care provider–focused websites providing update information (Table 4). Authorship was the most commonly missed criteria, with only 3 consumer-focused and 3 health care provider–focused websites properly attributing the authors of information, including their qualifications and affiliations. There was no significant difference in the mean number of criteria satisfied between consumer-focused and health care provider–focused websites.

The HONcode was only present on 2 websites (*NPS MedicineWise* and *Health Direct*; Table 2). No websites displayed an explicit conflict of interest statement (represented as part of the disclosure criteria); therefore, the existence of any actual or potential conflicts cannot be determined. *RACGP* provided a link to a conflict of interest management policy, which states that no one with a conflict of interest can be involved in the decision-making process. *NCSP Clinical Guidelines* provided a register of all interests, leaving it up to the end user to determine the presence and significance of any conflict.

Table 4. Number of included websites adhering to the Journal of the American Medical Association benchmark criteria.

JAMA[a] benchmark criteria	Websites adhering, n (%)		
	Consumer-focused websites	Health care provider–focused websites	All websites
Authorship	3 (30)	3 (60)	6 (40)
Attribution	6 (60)	2 (40)	8 (53)
Disclosure	7 (70)	4 (80)	11 (73)
Currency	8 (80)	4 (80)	12 (80)

[a] JAMA: Journal of the American Medical Association.

Website Design

All websites had consistent use of white space, separating paragraphs of information into smaller *chunks* of text. All websites used distinct and easy-to-read font styles and had clear subheadings in a text size larger than the main text. An adaptable font size was available in 2 of 15 websites (*Cancer Council Australia Main Site* and *Cancer Australia Cervical Cancer).* All 15 websites used a responsive web design. Of 15 websites, 3 (*NSW CSP, Cancer Council Australia Cervical Screening Consumer Site,* and *Jean Hailes)* used illustrations to reinforce key messages.

Simulation Questions

Of 15 websites, 3 (*NCSP Main Site, Cancer Council Australia Cervical Screening Consumer Site,* and *NSW CSP*) provided answers to all 5 simulation questions. The question most frequently answered by websites was question 5: “Who needs the new test?” *NPS MedicineWise, RACGP, National Cancer Screening Register,* and *Queensland CSP (B)* did not provide answers to any of the questions. The time taken to find the answers ranged from approximately 10 seconds to 2 min. The answers were found the fastest on *NSW CSP* (average time of 34 seconds per question) and the slowest on *NCSP Clinical Guidelines* (average time of 146 seconds per question).

Discussion

Principal Findings

This study is the first to evaluate web-based information available to consumers about the renewal of the Australian NCSP. Overall, the readability of consumer-focused websites providing information about the renewed NCSP was fairly difficult to read by the general population (ie, below the threshold of 60); however, there was high variability in scores. Overall, the evaluated websites demonstrated a high level of understandability. The *Cancer Council Australia Cervical Screening Consumer Site* scored highest for understandability assessed using the PEMAT. There was no significant correlation between reading ease and understandability, demonstrating that they measure different constructs. This is in line with previous studies and highlights the importance of using both measures to evaluate quality [17]. Most websites met some criteria to assess their credibility, and all websites demonstrated some thought had been given to the design of the websites. Of the 5 key questions that could be frequently searched by consumers, the following 3 websites provided answers to all questions: *NCSP Main Site, Cancer Council Australia Cervical Screening Consumer Site,* and *NSW CSP.* These findings demonstrate great variability in the readability, understandability, and credibility of websites available on the internet, which provide both consumers and health care providers with information about the changes to the NCSP.

Across all measures, the *Cancer Council Australia Cervical Screening Consumer Site* was judged to be the most accessible website overall. This site was considered the easiest to read based on the FRE score and had a reading level between grade 6 and 7 using both the FKGL and SMOG index, which is in line with recommendations from South Australia Health [11]. Furthermore, it scored highly on the understandability criterion of the PEMAT and was one of only three websites to answer all 5 simulation questions. This website was developed by Cancer Council Australia in response to the renewal of the NCSP and has a clear purpose to educate women about the changes to cervical screening and seeks to answer possible frequently answered questions. It provides comprehensive content about the changes in an accessible manner. Australian women with questions or concerns about the renewed NCSP may benefit from being directed to this website by their health care provider.

Strengths and Limitations

This study used validated tools and a combination of several measures to independently evaluate websites providing information about cervical screening. Objective readability measures provide limited data about the health literacy level required to understand information, so by combining this with the PEMAT, this enabled us to determine how well people of various backgrounds and levels of health literacy may be able to understand and act upon the provided health information. These tools were selected because the items were most relevant to the health information we were evaluating. Other available tools, such as DISCERN [28], have a number of items aimed at treatment-related information, which are not relevant for the content of cervical screening information. A limitation of our study is the small number of websites included in the analysis, which may have prevented seeing differences in readability between health care provider–focused and consumer-focused websites. However, the search term *cervical screening* returned over 50,000,000 results on Google alone, and as internet users rarely look past the first 3 results pages, it is likely that the websites included in our analysis would be those most commonly accessed by consumers [18]. Some of the measures we used are subjective (eg, PEMAT), but by having 2 independent evaluators score all 15 websites and a third evaluator score 40% (6/15) of websites, this will have helped ease any discrepancies, and the interrater reliability between the evaluators was rated as substantial. The websites in this study were only examined using their content and were not examined for accuracy of the information presented, as this was not an aim of the study. This study was limited to the evaluation of printable material on the websites; therefore, information delivered in video format was excluded.

Comparison With Previous Work

This study is particularly relevant in light of the renewed NCSP and the public misconception about the rationale behind the changes [5,6]. Our findings show that most consumer-focused web-based information on cervical screening in Australia is written at a reading level inaccessible to the general population. This is similar to previous studies that have found web-based health information to often be written above the recommended reading level [29]. This may be because the authors have insufficient awareness of health literacy considerations and readability [30].

In this study, websites with the lowest readability tended to be those targeted toward health care providers. Making the target audience of the website more explicit may help avoid potential confusion for consumers and the frustration of not having questions answered. Although this may be obvious for some websites (eg, *NCSP Clinical Guidelines*), health consumers should be able to determine at a glance the target audience of the information they are reading for all websites. This may minimize their exposure to inaccessible material. Sites aimed toward health care providers could provide links to guide consumers to reputable consumer-focused websites (eg, *NCSP Main Site* or *Cancer Council Australia Cervical Screening Consumer Site*).

The high understandability scores indicate that web-based cervical screening information can be understood relatively easily by consumers of diverse backgrounds and levels of health literacy [23] and demonstrate that the PEMAT and readability tools measure different constructs. Greater awareness of how the mode of delivery may affect consumers’ understanding of the health information could help to create and distribute better patient materials. The low actionability scores suggest that it may be difficult for consumers to act upon the information they read. Most websites failed to provide tangible tools or visual aids and lacked explicit instructions for the user to follow (eg, a tool to help consumers decide whether they are eligible for screening under the new guidelines). The general purpose of these websites is to inform and educate the reader about the changes rather than encouraging an action. Consequently, the actionability domain of the PEMAT may have limited applicability to this context. The large range in PEMAT scores may be partially explained by the fact that some websites were solely targeted toward health care providers. As such, the content on these websites (*NCSP Clinical Guidelines, RACGP,* and *NPS MedicineWise*) would not be considered *patient* education materials, and consequently, the PEMAT score may not be an accurate measure of quality. Most existing literature on web-based health information is related to disease and treatment information, where there may be more explicit actions to take than for cervical screening (eg, different treatment options for a chronic disease). This may, in part, contribute to the low actionability scores for our websites, but because of actionability not being an aim of these websites, less weight was given to these scores.

Noting the presence of the HONcode logo is a simple way for consumers to recognize that the information they are reading is trustworthy, but because HONcode certification is a voluntary registration process, it is possible that websites may comply with the principles despite not being certified [10]. As most of the included websites are federal or state government owned, they would likely be instantly recognized as trustworthy and credible sources of information; therefore, there may be no added benefit of HONcode certification. It is unclear how recognizable the HONcode is among both consumers and website developers. The only 2 websites displaying the HONcode are government-funded not-for-profit organizations (*Health Direct* and *NPS Medicine Wise*). As these sites may not be as easily recognized as reputable sources, HONcode certification may be more valuable; however, there is little value if a consumer does not know what it is or how to interpret it.

Most websites failed to provide adequate detail about the authors of the information or to provide sources and references for information. This information was seldom displayed prominently, often located on the *about us* or *disclaimer* pages, which are rarely accessed by consumers [18]. Sources that were provided were rarely primary research and often linked to related websites, raising the potential of consumers to be exposed to inaccurate, inaccessible, or unreliable content if the linked websites are of poor quality or lack credibility. It could be argued, however, that the omission of primary sources on consumer-focused websites may be a deliberate decision by the website developer. Primary research papers are often behind paywalls or written at a reading level inaccessible to the general public. This has the potential to cause greater confusion and may offer little benefit to the reader. Furthermore, consumers generally judge the credibility of websites based on the hosting organization (ie, whether it is from a reputable and recognizable organization such as the Australian Government or Cancer Council Australia), language, and professional-looking design [18]. Web-based information can be biased, with little to no scientific evidence [18]. In addition to providing credit to the original source, appropriate attribution can help the end user evaluate the quality and ensure the trustworthiness of information on the web. This leads to a discussion of whether there should be mandatory criteria to follow when setting up websites to ensure their credibility. Furthermore, it raises the question of whether details about sources, authors, and conflicts of interest can be displayed in a way that is more accessible and meaningful to health consumers.

The 3 websites that satisfied all 4 JAMA benchmarks were targeted specifically toward health care providers. If websites targeted toward health consumers lack the aforementioned principles, it raises potential issues. The general population, particularly those of lower health literacy, may be less equipped than health care providers to evaluate and discredit websites. It is critical that consumer-focused websites are credible, but it may be necessary for a tradeoff between accessibility, readability, understandability, and credibility. This area requires further research to minimize this tradeoff, as each component contributes toward consumers’ experience of accessing health information on the web. Appropriate attribution and references are a particularly important credibility indicator of health care provider–focused websites; however, they may be considered less relevant for consumer-focused sites. Existing credibility assessment tools (such as the JAMA benchmark criteria and the HONcode criteria) are limited in their ability to evaluate consumer-focused websites. The development of a new credibility framework for consumer-focused websites could mitigate this limitation. Taking into consideration other factors that influence credibility, such as placing more weight on the hosting organization, could enable consumers to discredit unreliable information without reducing the overall quality.

Owing to an increasing interest in the literature about conflicts of interest, it was interesting to find that none of the websites in this study explicitly disclosed a conflict of interest. It is important to note that the absence of a conflict of interest statement is not equivalent to having no conflict. There is increasing research into the management of conflicts of interest and whether or not disclosure is the most appropriate strategy [31]. Disclosing conflicts merely indicates the possibility of a bias; however, it does not provide guidance for resolving it or how the conflict was managed. For the consumer, a lack of disclosure may hinder their ability to mitigate any potential risk of bias, as conflicts of interest may impact decision making [31].

Simulation questions were helpful in establishing whether these websites explained the changes to the CSP. The majority of these websites did not answer the questions in detail, raising concerns about whether consumers reading the information would be fully informed about the changes. A total of 3 websites provided answers to all the questions (*NCSP Main Site*, *Cancer Council Australia Cervical Screening Consumer Site*, and *NSW CSP*), with the time taken to find answers to these questions being the quickest for the *NCSP Main Site.* The websites that either did not provide answers to the simulation questions or it took a long time to find the answers were those aimed toward health care providers, further highlighting the need for the target audience of the website to be made more explicit.

Finally, all websites were designed to improve the accessibility of information through the use of white space and appropriate font style. The layout and design of materials can be a contributing factor to how consumers comprehend the information or how accessible the information is [22]. Only 2 websites showed adaptable font size, which may cause potential difficulties for those with impaired sight. An important consideration when designing websites now is also about how the interface looks on mobile phones, given the increasing number of consumers who will view health information on their mobile phones [7]. It was therefore reassuring to observe that all websites used responsive web design. The use of illustrations can reinforce messages and aid interpretation of written text or they can distract away from the purpose [26]. The *NSW CSP* and *Cancer Council Australia Cervical Screening Consumer Site* use photographs of culturally and demographically diverse women, which may help consumers understand that the information is relevant to them.

Conclusions

The findings from this study can help health care providers direct their patients toward websites that have information about the renewal of the CSP, which is easy to read and has high understandability (eg, *Cancer Council Australia Cervical Screening Consumer Site*). Encouragingly, web-based information about the renewed CSP is generally of good quality. There is a need to promote increased awareness of the importance of web-based health information that is credible, user friendly, and easily understood by people with wide levels of health literacy. Web-based information should be produced with particular consideration to people with low health literacy, with those responsible for creating these websites to be accountable for ensuring that this information is accessible, accurate, and credible. There may be potential for web-based health information to be improved through the introduction of mandatory criteria to ensure the credibility and quality of consumer-focused websites.

Authors' Contributions

RD conceived and designed the study and contributed to the drafting of the manuscript. AT and ST acquired the data set, analyzed the data, and contributed to the manuscript. OM drafted the manuscript and contributed to the analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of website evaluation measures.
[PDF File (Adobe PDF File), 95 KB - jmir_v22i6e16701_app1.pdf]

Multimedia Appendix 2
National Institute of Adult Continuing Education tool, Readability: producing clear written materials for a range of readers.
[PDF File (Adobe PDF File), 376 KB - jmir_v22i6e16701_app2.pdf]

Multimedia Appendix 3
Website URLs of all identified websites by search order.
[PDF File (Adobe PDF File), 140 KB - jmir_v22i6e16701_app3.pdf]

Multimedia Appendix 4
Website abbreviations, organisation and target audience of included websites.
[PDF File (Adobe PDF File), 90 KB - jmir_v22i6e16701_app4.pdf]

References

1. Australian Institute of Health and Welfare. 2019. National Cervical Screening Program Monitoring Report 2019 URL: https://www.aihw.gov.au/getmedia/fcacac12-cd05-4325-88bc-5529a61b53f3/aihw-can-132.pdf.aspx?inline=true [accessed 2019-09-10]
2. Australian Institute of Health and Welfare. 2019. Cervical Screening in Australia URL: https://www.aihw.gov.au/reports/cancer-screening/cervical-screening-in-australia-2019/contents/table-of-contents [accessed 2019-09-10]
3. Cancer Screening. 2013. National Cervical Screening Program Renewal: Evidence Review URL: http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/E6A211A6FFC29E2CCA257CED007FB678/%24File/Review%20of%20Evidence%20notated%2013.06.14.pdf [accessed 2019-10-10]
4. Lew JB, Simms KT, Smith MA, Hall M, Kang YJ, Xu XM, et al. Primary HPV testing versus cytology-based cervical screening in women in Australia vaccinated for HPV and unvaccinated: effectiveness and economic assessment for the national cervical screening program. Lancet Public Health 2017 Feb;2(2):e96-107 [FREE Full text] [doi: 10.1016/S2468-2667(17)30007-5] [Medline: 29253402]
5. Obermair HM, Dodd RH, Bonner C, Jansen J, McCaffery K. 'It has saved thousands of lives, so why change it?' Content analysis of objections to cervical screening programme changes in Australia. BMJ Open 2018 Feb 13;8(2):e019171 [FREE Full text] [doi: 10.1136/bmjopen-2017-019171] [Medline: 29440214]
6. Dodd RH, Obermair HM, McCaffery KJ. A thematic analysis of attitudes toward changes to cervical screening in Australia. JMIR Cancer 2019 Apr 11;5(1):e12307 [FREE Full text] [doi: 10.2196/12307] [Medline: 30973340]
7. Cheng C, Dunn M. How well are health information websites displayed on mobile phones? Implications for the readability of health information. Health Promot J Austr 2017 Mar;28(1):15-20. [doi: 10.1071/HE15127] [Medline: 27250916]
8. Cocco AM, Zordan R, Taylor DM, Weiland TJ, Dilley SJ, Kant J, et al. Dr Google in the ED: searching for online health information by adult emergency department patients. Med J Aust 2018 Aug 20;209(8):342-347. [doi: 10.5694/mja17.00889]
9. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 2;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
10. The Australian Curriculum, Assessment and Reporting Authority. 2011. National Report on Schooling in Australia 2009 URL: https://www.acara.edu.au/reporting/national-report-on-schooling-in-australia/nrosia2009 [accessed 2020-02-13]

11. Australian Commission on Safety and Quality in Health Care. 2013. Health Literacy: Taking Action to Improve Safety and Quality URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/Health-Literacy-Taking-action-to-improve-safety-and-quality.pdf [accessed 2020-05-04]
12. Grabeel KL, Russomanno J, Oelschlegel S, Tester E, Heidel RE. Computerized versus hand-scored health literacy tools: a comparison of simple measure of gobbledygook (SMOG) and Flesch-Kincaid in printed patient education materials. J Med Libr Assoc 2018 Jan;106(1):38-45 [FREE Full text] [doi: 10.5195/jmla.2018.262] [Medline: 29339932]
13. Farnsworth M. Differences in perceived difficulty in print and online patient education materials. Perm J 2014;18(4):45-50 [FREE Full text] [doi: 10.7812/TPP/14-008] [Medline: 25662526]
14. Nghiem A, Mahmoud Y, Som R. Evaluating the quality of internet information for breast cancer. Breast 2016 Feb;25:34-37. [doi: 10.1016/j.breast.2015.10.001] [Medline: 26547835]
15. Lipari M, Berlie H, Saleh Y, Hang P, Moser L. Understandability, actionability, and readability of online patient education materials about diabetes mellitus. Am J Health Syst Pharm 2019 Jan 25;76(3):182-186. [doi: 10.1093/ajhp/zxy021] [Medline: 31408087]
16. Kirby PL, Reynolds KA, Walker JR, Furer P, Pryor TA. Evaluating the quality of perinatal anxiety information available online. Arch Womens Ment Health 2018 Dec;21(6):813-820. [doi: 10.1007/s00737-018-0875-5] [Medline: 29931445]
17. Aguirre PE, Coelho MM, Rios D, Machado MA, Cruvinel AF, Cruvinel T. Evaluating the dental caries-related information on Brazilian websites: qualitative study. J Med Internet Res 2017 Dec 13;19(12):e415 [FREE Full text] [doi: 10.2196/jmir.7681] [Medline: 29237585]
18. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 9;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
19. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
20. McLaughlin GH. SMOG grading: a new readability formula. J Read 1969;12(8):639-646 [FREE Full text]
21. Beaunoyer E, Arsenault M, Lomanowska AM, Guitton MJ. Understanding online health information: evaluation, tools, and strategies. Patient Educ Couns 2017 Feb;100(2):183-189. [doi: 10.1016/j.pec.2016.08.028] [Medline: 27595436]
22. Shedlosky-Shoemaker R, Sturm AC, Saleem M, Kelly KM. Tools for assessing readability and quality of health-related web sites. J Genet Couns 2009 Feb;18(1):49-59. [doi: 10.1007/s10897-008-9181-0] [Medline: 18937063]
23. Shoemaker SJ, Wolf MS, Brach C. Development of the patient education materials assessment tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014 Sep;96(3):395-403 [FREE Full text] [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
24. Boyer C, Selby M, Scherrer J, Appel R. The health on the net code of conduct for medical and health websites. Comput Biol Med 1998 Sep;28(5):603-610. [doi: 10.1016/s0010-4825(98)00037-7] [Medline: 9861515]
25. Aday LA. Benchmarks of fairness for health care reform. J Am Med Assoc 1997 Jan 22;277(4):347-350 [FREE Full text] [doi: 10.1001/jama.1997.03540280085042]
26. Learning and Work Institute. 2009. Readability: How to Produce Clear Written Materials for a Range of Readers URL: https://www.learningandwork.org.uk/resource/readability/ [accessed 2019-04-06]
27. WebFX. Readability Test Tool URL: https://www.webfx.com/tools/read-able/ [accessed 2019-04-06]
28. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
29. Cheng C, Dunn M. Health literacy and the Internet: a study on the readability of Australian online health information. Aust N Z J Public Health 2015 Aug;39(4):309-314. [doi: 10.1111/1753-6405.12341] [Medline: 25716142]
30. Boztas N, Omur D, Ozbılgın S, Altuntas G, Piskin E, Ozkardesler S, et al. Readability of internet-sourced patient education material related to 'labour analgesia'. Medicine (Baltimore) 2017 Nov;96(45):e8526 [FREE Full text] [doi: 10.1097/MD.0000000000008526] [Medline: 29137057]
31. Dunn AG, Coiera E, Mandl KD, Bourgeois FT. Conflict of interest disclosure in biomedical research: a review of current practices, biases, and the role of public registries in improving transparency. Res Integr Peer Rev 2016;1(1) [FREE Full text] [doi: 10.1186/s41073-016-0006-7] [Medline: 27158530]

Abbreviations

CSP: Cervical Screening Program
FKGL: Flesch-Kincaid Grade level
FRE: Flesch Reading Ease
HONcode: Health on the Net Foundation code of conduct
HPV: human papillomavirus
JAMA: Journal of the American Medical Association
NCSP: National Cervical Screening Program

NIACE: National Institute of Adult Continuing Education
PEMAT: Patient Education Materials Assessment tool
RACGP: Royal Australian College of General Practice
SMOG: Simple Measure of Gobbledygook index

Edited by G Eysenbach; submitted 06.11.19; peer-reviewed by M Smith, T Cruvinel, S Bidmon; comments to author 23.12.19; revised version received 13.02.20; accepted 09.04.20; published 26.06.20.

Please cite as:
Mac OA, Thayre A, Tan S, Dodd RH
Web-Based Health Information Following the Renewal of the Cervical Screening Program in Australia: Evaluation of Readability, Understandability, and Credibility
J Med Internet Res 2020;22(6):e16701
URL: http://www.jmir.org/2020/6/e16701/
doi:10.2196/16701
PMID:32442134

Original Paper

Cigarette Packs With URLs Leading to Tobacco Company Websites: Content Analysis

Caitlin Victoria Weiger[1], MHS; Katherine Smith[1], PhD; Amy Y Hong[2], MHS; Joanna E Cohen[1], PhD

[1]Institute for Global Tobacco Control, Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Caitlin Victoria Weiger, MHS
Institute for Global Tobacco Control
Department of Health, Behavior and Society
Johns Hopkins Bloomberg School of Public Health
2213 McElderry St
4th Floor
Baltimore, MD, 21205
United States
Phone: 1 410 614 5378
Fax: 1 410 614 1003
Email: cweiger2@jhmi.edu

Abstract

Background: Tobacco companies include on the packaging of their products URLs directing consumers to websites that contain protobacco messages. Online media tend to be underregulated and provide the industry with an opportunity to present users with protobacco communication.

Objective: The objective of our study was to document the content of websites that were advertised on tobacco packs in 14 low- and middle-income countries.

Methods: We purchased tobacco packs from 14 low- and middle-income countries in 2013 and examined them for the presence of URLs. We visited unique URLs on multiple occasions between October 1, 2016 and August 9, 2017. We developed a coding checklist and used it to conduct a content analysis of active corporate websites to identify types of protobacco communication. The coding checklist included the presence of regulatory controls and warnings, engagement strategies, marketing appeals (eg, description of product popularity, luxury/quality, taste), corporate social responsibility programs, and image management. We coded brand websites separately and also described social media and other website types.

Results: We identified 89 unique URLs, of which 54 were active during the search period. We assessed 26 corporate websites, 21 brand websites, 2 nontobacco websites, and 5 social media pages. We excluded 2 corporate websites and 14 brand websites due to limited accessible content or incomplete content. Corporate social responsibility was discussed on all corporate websites, and marketing appeals were also common. Corporate websites were also more likely to include more nonspecific (12/24, 50%) than specific (7/24, 29%) health warnings. Promotions (6/7, 86%) and sociability appeals (3/7, 43%) were common on brand websites. The small number of social media webpages in our sample used gendered marketing.

Conclusions: URLs appearing on tobacco packs direct consumers to websites where users are exposed to marketing that highlights the “positive” contributions of tobacco companies on corporate websites, and extensive promotions and marketing appeals on brand websites and social media pages. It is essential that marketing regulations become more comprehensive and ban all protobacco communication, a policy that is in line with articles 5.3 and 13 of the World Health Organization Framework Convention on Tobacco Control. For countries that already ban internet tobacco advertising, enforcement efforts should be strengthened. Tobacco companies’ use of URLs on packs may also be compelling for plain packaging advocacy, where all branding is removed from the pack and large graphic health warning labels are the only communication on the tobacco packaging. Future research should consider including tobacco websites in marketing surveillance.

(J Med Internet Res 2020;22(6):e15160) doi:10.2196/15160

KEYWORDS

advertising; developing countries; product packaging; product labeling; tobacco industry

Introduction

Background

Tobacco is the most dangerous consumer product in the world, and tobacco product marketing has been recognized as a cause of tobacco use initiation and failed quit attempts for over a decade [1-10]. The internet provides an additional platform for protobacco communication via tobacco industry websites. The importance of these websites is highlighted by the millions of dollars the industry spends on them; for example, US tobacco companies have been increasing spending on internet marketing since the early 2000s, spending US $25.1 million in 2017 on company websites [11,12]. The continually evolving nature of the internet and tobacco industry marketing strategies makes these websites a public health surveillance priority [13].

Tobacco companies use these websites to promote both their product branding and their corporate image. Given the broad public awareness of the harms of smoking and their deceptive tactics, tobacco companies have taken steps to try to counteract negative perceptions by engaging in activities that they describe as serving to build credibility [14]. The industry touts green business practices, publicizes support for what they see as reasonable (one might imagine, restricted) approaches to regulation, acknowledges some harms caused by smoking, and promotes the future of “reduced-harm” products in order to manage their corporate image and normalize tobacco companies’ practices [14-16]. Although normalization of the industry is itself harmful, these activities additionally enhance industry access to policy makers, influence the credibility of industry-proposed policy alternatives, and weaken opposing political constituencies [17,18]. While the industry formerly relied on traditional media and opinion leaders to spread these messages, they now also incorporate online content, such as company websites [16,19-22]. Tobacco companies print URLs to their websites on their product packaging, presumably increasing exposure. URLs on packs link consumers to much more protobacco communication than can otherwise be contained on the pack.

There are several types of tobacco websites, which tend to fall into 3 main categories: corporate [20], brand [23,24], and social media [25]. Corporate websites represent the company as a formal business and present their corporate image to the public [26,27]. Brand websites market specific products through “interactivity, multidirectional communication and meaningful engagement” [24]. Social media pages can be associated with either a company or a brand, or may be the product of consumer-generated content [28]. Tobacco companies can make “pages” or accounts on unaffiliated social networking platforms that people can like or follow. As an example, Dunhill, a brand of British American Tobacco cigarettes, had a page on Facebook in 2009 to promote Dunhill events, including parties at nightclubs, where free Dunhill cigarettes were distributed [29]. Nontobacco websites, websites not visibly associated with a tobacco company, are also sometimes listed on tobacco packs.

Despite its potential to influence both consumer and policy-maker perceptions of the tobacco industry, online tobacco marketing is still relatively new and underregulated [28,30]. Legislation on tobacco marketing rarely mentions internet marketing specifically, leaving the status of what types of marketing are allowed online up to interpretation (Multimedia Appendix 1). It is also unclear whether corporate marketing for specific companies that promotes the company rather than its products is allowed under advertising and marketing bans.

Tobacco control researchers have already started to document activity on tobacco websites. Previous studies on corporate websites have focused on corporate social responsibility (CSR), the provision of limited health information, and image management [20,26,31,32]. Brand websites were found to tout interactive games, contests, event information, and coupons, and even to allow users to select their own music to enjoy while browsing [22,33-36].

The literature on tobacco internet marketing is, however, somewhat limited. Content analyses of corporate websites have focused on samples of US websites or only the websites of the largest multinational companies [20,26,34]. Research has not investigated websites, even for multinational companies, in low- and middle-income countries (LMICs), where 80% of daily smokers live and where tobacco companies are focusing efforts to expand the cigarette market [37,38]. Research on brand websites also tends to be older and limited to a case study approach, and the constantly changing nature of the internet may make some of those findings less relevant today than when they were originally published [22,23,39,40].

Objective

To help fill these gaps, we conducted a quantitative content analysis of corporate tobacco and brand tobacco websites and described nontobacco websites and social media pages associated with URLs that we found on tobacco packs purchased in 14 LMICs. Our objective was to assess the marketing strategies to which tobacco companies are linking their consumers. This descriptive analysis may inform marketing regulations and further support packaging restrictions and plain packaging initiatives.

Methods

Sample

As part of the Tobacco Pack Surveillance System (TPackSS) study [41], 3307 cigarette, kretek (clove cigarettes common in Indonesia) [42], and bidi (hand-rolled cigarettes common in India and Southeast Asia) [43] country-unique packs were systematically purchased from 14 LMICs in 2013 [41]. We recorded each specific URL using pack photos available from TPackSS [44] and in-person examination of the pack when the URL was not legible from photos (n=56). URLs were located on all sides of the pack (front, back, sides, top, and bottom), as well as on the inside of the packs (inside lid and on paper inserts included in the packs). We entered URLs into an internet search to locate the website, and downloaded and saved screenshots of the home page and each main tab of the menu that was one click away from the home page. While this approach did not allow for the assessment of all content on each website, we established a limit, as it was not feasible to follow every link to every page. We visited websites multiple times between October 1, 2016 and August 9, 2017 to maximize the number of active websites in our sample, as it was common for websites to go through cycles of inactivity and activity.

Procedure

If an English-language version of the website was not provided, we asked a translator to write the English translation of the text on the saved PDFs of website images. Translators were then present (either physically or via Skype; Microsoft Corporation) while coders coded these PDFs in the event that coders had questions about the meaning of the translation. We used codes from a content analysis of Russian e-cigarette websites [45] and common features found on US tobacco websites during field testing to develop a codebook to quantitatively inventory website content. Categories were engagement strategies, marketing appeals, regulatory controls and warnings, and CSR and image management. We considered graphic content only for codes related to flavor and marketing appeals that could be easily and reliably identified (ie, luxury/quality, sociability, masculinity, and femininity appeals). We programmed the coding checklist into Qualtrics software version 05.2017 to 09.2017 (Qualtrics, LLC) to facilitate coding (Multimedia Appendix 2).

Due to high interrater reliability assessed after coding the first 20 websites, not all of the remaining websites were double coded, resulting in 81% (25/31) of quantitatively coded (ie, corporate and brand) websites being double coded. We coded the websites' home page and each main tab from the website menu for presence versus absence of each feature. We assessed interrater reliability using percentage agreement, as well as a method to account for the low prevalence of some marketing strategies (prevalence-adjusted and bias-adjusted kappa) [46-48]. Overall percentage agreement was .95 and prevalence-adjusted and bias-adjusted kappa was .89. Discrepancies were reviewed and reconciled via in-person discussion between the coders, with clarifications made to the codebook as needed.

Analysis

Data were exported from Qualtrics into an Excel spreadsheet version 16.6 (Microsoft Corporation) and analyzed using Stata version 15 (StataCorp LP). We calculated the proportion of websites whose content reflected each of the codes.

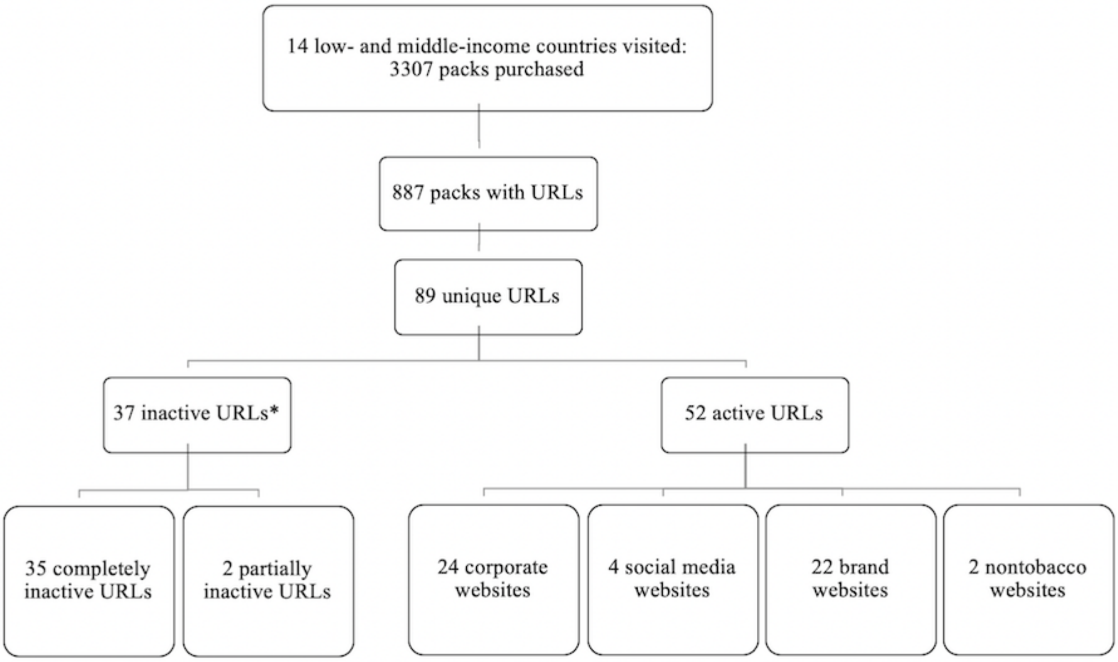
We assessed brand websites using the same coding checklist used for corporate websites, although we present the results separately due to the limited amount of content visible without registration. We briefly described the social media pages and nontobacco websites, since their characteristics were not adequately captured by the codebook.

Results

Website Characteristics

We noted the presence and location of an URL on 880 packs (857 cigarette packs and 23 kretek packs). Of these, 7 packs included 2 URLs. There were 54 active and unique websites: 26 were corporate websites, 21 were brand websites, 5 were social media websites, and 2 were nontobacco websites. We excluded 2 corporate websites because not all website pages were active (see Figure 1 for details). All brand websites but 1 required the user to register in order to see complete website content. Registration required local addresses, verification of country citizenship, and verification that you were a tobacco product user. We excluded 14 brand websites with no content visible prior to registration, as we could not breech the terms and conditions of the websites per university requirements.

Figure 1. Breakdown of coded websites. *37 websites considered inactive for the following reasons: the page could not be found (n=17), the page took too long to respond (n=7), the site was closed for construction (n=5), the domain name had expired (n=4), some parts of the page were inactive (n=2), the company had been sold (n=1), or access was not allowed (n=1).



The 38 websites that were accessible for analysis were from 560 packs, or 63.6% of all 880 purchased packs with URLs. These packs were from 13 countries, and the websites were available in over 20 languages, counting all languages available on each webpage (Multimedia Appendix 3). We found the 24 active corporate websites included in this analysis on 486 packs purchased in 13 countries, or 14.70% of all 3307 purchased packs and 55.2% of all 880 purchased packs with URLs. We categorized 8 websites as multinational and 16 as nonmultinational. In line with work by others, we considered the 4 leading multinational tobacco companies, Philip Morris International, British American Tobacco, Japan Tobacco International, and Imperial Tobacco [49], to be multinational. Other corporate websites were from state-owned and other regional or national companies. The 7 active brand website URLs with at least some content visible were on 52 packs from 6 countries, the 5 social media webpage URLs were on 19 packs from 3 countries, and the 2 nontobacco website URLs were on 3 packs from 2 countries (Multimedia Appendix 3).

Corporate Websites

Regulatory Controls and Warnings

Of the 24 corporate websites, 14 (58%) had statements regarding the legal age for product purchase or website use (Table 1). We found pop-up windows asking whether the user was over the age of 18 years (known as click-through verifications) on 3 (13%) websites, but no website had a specific verification requirement (eg, exact date of birth). Of the 24 websites, 7 (29%) had specific health warnings (eg, from the Turkish Imperial Tobacco website: “Smoking is a cause of serious diseases in smokers, including lung cancer, heart disease and emphysema”), whereas 12 (50%) had only general acknowledgements of the health risks of smoking (eg, from the Tutun company website: “Smoking is negatively impacting your health”).

Table 1. Codebook used for corporate (n=24) websites and proportion of corporate websites by the presence of each code.

Domain	Code prevalence on corporate websites, n (%)
Regulatory controls and warnings	
Access restrictions	3 (13)
Age-restricted product usage statements	14 (58)
General health warning	12 (50)
Specific health warning	7 (29)
Engagement strategies	
Link to additional corporate website (eg, parent website)	7 (29)
Link to a brand website	2 (8)
Link to a social media page	14 (58)
Link to another type of webpage	15 (63)
Option to download an app	2 (8)
Email alerts	3 (13)
Contests	1 (4)
Video	9 (38)
Scrolling images	15 (63)
Marketing appeals	
Product popularity	13 (54)
Luxury/quality	21 (88)
Femininity	1 (4)
Masculinity	0 (0)
Taste/sensation	10 (42)
Flavor other than menthol	6 (25)
Menthol flavor	3 (13)
International appeal	19 (79)
National appeal	19 (79)
US appeal	6 (25)
Trendsetter	1 (4)
Tradition	16 (67)
Light	2 (8)
Low tar	3 (13)
Other harm reduction term	5 (21)
CSR[a] and image management	
General mention of CSR	17 (71)
Educational support	8 (33)
Investment in community	10 (42)
Harm reduction technology	5 (21)
Provider of jobs	14 (58)
Employee care/support	16 (67)
Youth smoking prevention program	3 (13)
General environmental claim	11 (46)
Specific environmental claim	6 (25)
Performance or CSR awards	13 (54)

Domain	Code prevalence on corporate websites, n (%)
Description of consumers as adults	9 (38)
Messaging about informed consumers	6 (25)
Support for tobacco control legislation	9 (38)
Transparency	9 (38)
Innovative	19 (79)

[a] CSR: corporate social responsibility

Engagement Strategies

A few corporate websites offered opportunities for further engagement with their company by offering mobile apps (n=2, 8%) and prompting users to sign up for email alerts (n=3, 13%) (Table 1). British American Tobacco's website had an app for investors to track company performance. British American Tobacco and Imperial Tobacco both offered email alerts for company media coverage. Some websites also linked to other corporate websites, including the parent company (n=7, 29%), brand (n=2, 8%), and social media websites and webpages (n=14, 58%).

Marketing Appeals

Every corporate website contained at least one marketing appeal for their products or company (Table 1). Luxury and quality appeals were the most common (n=21, 88%), such as British American Tobacco's description of its "superior quality next generation products." National appeals (n=19, 79%) and appeals to company tradition or history (n=16, 67%) were common. Hongta Tobacco Group Co., a Chinese tobacco company, appealed to company tradition by dedicating a page of its website to the history of the company from its founding in 1956 to its "courageous" introduction of advanced foreign technology. We found popularity appeals, such as Altria Group's boasting of their ownership of Philip Morris USA ("Leading cigarette manufacturing in the U.S. and has been for more than 30 years") and that of U.S. Smokeless Tobacco Co. (the "World's leading producer of moist smokeless tobacco"), on 13 (54%) websites. Appeals to the taste or sensation of tobacco were present on 10 (42%) corporate websites. For instance, the Djarum website described how the "exotic flavors of Djarum kretek has captured the imagination of a growing international audience" [50], using both taste or sensation and international appeals. We also observed international appeals (n=19, 79%), including international appeals mentioning American culture or values (n=6, 25%).

Corporate Social Responsibility and Image Management

All 24 corporate websites mentioned CSR or a specific CSR program at least once. Examples of CSR were providing scholarships or other academic support (n=8, 33%), caring for employees and providing support (n=16, 67%), being a provider of jobs (n=14, 58%), and using or investing in harm reduction technology (n=5, 21%). The website for British American Tobacco Russia stated that they were one of the best employers in Russia and claimed to offer employees "all possibilities for professional and career growth." Hongta publicized several educational initiatives on their websites, including a ¥4 million (about US $565,000) donation to a middle school and the completion of a dormitory building project for another school. Youth smoking prevention programs were referenced on 3 (13%) websites: 2 British American Tobacco websites and Hongta's websites. A total of 10 (42%) websites discussed support for communities, including Japan Tobacco International via their ARISE program that is "bringing an end to child labor by working with tobacco farming communities to create new traditions." A total of 11 (46%) websites contained general references to environmental sustainability, such as British American Tobacco's "comprehensive environmental management program" [51]. A further 6 (25%) websites provided specific examples of sustainable or environmentally friendly practices, including Shaanxi's "energy-saving emission reduction program." Company corporate image was portrayed via mention of company transparency (n=9, 38%), descriptions of the company as innovative (n=19, 79%), discussion of the importance of consumer education (n=6, 25%), and mention of awards won for performance or CSR activities (n=13, 54%). The Russian company Donskoy Tabak endorsed their "principles of openness and transparency" and showcased their awards for "Quality and Technological Effectiveness." In addition, 9 (38%) websites expressed support for tobacco control legislation or emphasized compliance with such legislation. For example, Imperial Tobacco stated that "We recognize the controversial nature of our product and believe it is important to have it produced and handled by a responsible business. We both comply with and exceed current regulatory standards as we apply our own corporate standards."

Brand Websites

Although we coded only 7 brand websites because the other 14 brand websites in the sample set conditions for access incompatible with our university research standards, there are a few observations of note. Of these 7 websites, 6 limited the content visible without registration, while 1 was totally accessible without registration. Despite the limited content visible without registration, we observed promotions such as games, contests, and sponsored events (n=6, 86%), and youth appeals (n=2, 29%). A total of 3 (43%) brands websites included sociability appeals, which were illustrated on Bond Street's website through images of a small group of attractive young adults playing a guitar by the water. We noted national appeals, including a Marlboro-sponsored event called M-City, which attempted to express the culture of Kiev through art by inviting artists and musicians to the city, on 5 (71%) of the websites. Images of cities at night were common, and some websites included music. Most websites (n=6, 86%) used video or scrolling images, such as scenes of dance parties and cities at night on the Clas Mild websites and videos of "life hacks" on the L&M website.

Social Media

Of the 5 social media pages in this sample, 4 required that users request to join the group or register to see all content. Our requests to join were not accepted, so only a small amount of content was visible. The 2 social media pages for Kiss cigarettes were both characterized by femininity appeals, with flowers and text describing the page as a group for "modern girls." The Style Super Slim Tumblr-style social media page included posts on makeup, nail art, fashion, decorating, and baking. Another social media page for 21 Bek, however, used explicit masculine appeals, calling the page a group for "real men." Black Xperience also used masculine messaging and was a Tumblr-style website with images of cars, motorcycles, attractive women, and something called the "Black Innovation Challenge."

Websites Not Related to Tobacco

Two URLs found on packs led to websites that were not for tobacco companies or for tobacco products. One website was for the Programme for the Endorsement of Forest Certification, which described an agenda of caring for the world's forests and promoting sustainable forest management. Information on sustainability was available, and there was an option to sign up for newsletters from the organization. The other website was for the Japanese Ministry of Health, and the URL is part of the pack warning label required in Japan, although the 2 packs with this URL in the warning were purchased in Turkey. The Japanese Ministry of Health website contained information about the health hazards posed by tobacco, tobacco cessation, a passive smoke-free society movement, and a discussion of the Japanese Ministry's involvement with the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC).

Discussion

Principal Findings

URLs on tobacco packs purchased in LMICs led to corporate, brand, social media, and other websites that linked consumers to protobacco communication, including marketing for tobacco products and the companies that produce them. Several findings regarding the low prevalence of specific health warnings, use of CSR and other image management techniques, and use of marketing and engagement strategies highlight the potential of tobacco websites to communicate positive messages about the tobacco industry and tobacco products to users. These findings highlight the importance of tobacco websites in the tobacco marketing landscape in LMICs. There is a need for focused research on tobacco marketing in LMICs, as this is where the majority of tobacco users reside and where the industry has increasingly marketed its deadly products [37,38]. Tobacco websites should be explicitly included in marketing restrictions and regulations to protect the public from protobacco communication.

Corporate websites largely failed to communicate many of the harmful effects of smoking to website users. While half of the websites in this sample had a health warning, most of these warnings described vague risks associated with using tobacco, and risk of death and specific diseases were uncommon. However, having a health warning, even a nonspecific one, could be a way for a company to mitigate its liability [52] while presenting a more positive and responsible face to the public [32].

The emphasis of CSR programs on corporate websites was also of particular concern, as these programs can increase perceptions that the company is socially responsible, liking for the company, and positive associations with the company's products [53]. CSR has been used as a reputation and impression management tool to gain access to decision makers and increase tobacco company influence [17], and has been found to directly interfere with attempts to denormalize the tobacco industry and its products [54]. The very act of selling cigarettes, an addictive product that kills half its long-term users, is antithetical to CSR [55]. Additionally, corporate websites describe themselves as transparent, responsible, innovative companies that acknowledge the importance of consumer education, support tobacco control legislation, and even receive awards for their performance and CSR efforts. These strategies are used to promote an image of tobacco companies as socially responsible. This image serves their own monetary gain rather than the well-being of current and future tobacco consumers, who would be better served by the tobacco industry abandoning advertising and promotion of their products, and not opposing government regulation that aims to reduce tobacco-caused death and disease [56]. Articles 5.3 and 13 of the FCTC further recommend that parties mitigate the effect of "commercial and other vested interests of the tobacco industry" by denormalizing or regulating tobacco industry CSR or by banning CSR entirely by defining it as a form of sponsorship [57,58].

Corporate websites also promoted youth smoking prevention programs. Industry-funded youth smoking prevention advertising has been associated with increased intentions to smoke [59]. The WHO calls on FCTC parties to decline any industry involvement in tobacco use prevention [60], but our research demonstrated that the industry is still involved in tobacco use prevention online. Efforts should be made to keep the tobacco industry out of all tobacco control initiatives, across all media and both online and offline.

While corporate websites did not have contests, coupons, and games that have been documented on brand websites [22,33,34], corporate websites used other types of engagement strategies such as apps and email alerts. Corporate websites also included additional links to other corporate tobacco websites, brand websites, and social media pages, connecting users with even more protobacco content. This is in line with other recent findings on brand websites, which similarly found that all assessed brand websites contained at least one link to a social media page [34,61]. Tobacco companies used each communication platform left open to them to link to other platforms to further spread their messages.

Although our assessment of brand websites was limited by registration requirements, the high prevalence of promotional material visible even from the registration page is cause for concern. Exposure to promotions is associated with youth and young adult initiation, as well as lower likelihood of smoking cessation [1,62-65]. The use of music, both playing in the background and as part of some promotions, is also concerning given that the industry has previously employed music as a way to target youth and racial and ethnic minorities [66]. Brand websites frequently used sociability appeals and were often portrayed through images of attractive young people gathering with friends. Such marketing may be more likely to reach young people online, where they are increasingly spending more of their time. The limited amount of information we were able to observe on social media webpages indicated that this type of website may be more focused on marketing to specific subgroups, as evidenced by the explicitly gendered appeals found on parts of the websites that we were able to access.

URLs on cigarette packs can link consumers with corporate and brand marketing that in turn may undermine restrictions imposed on the physical marketing environment. Monitoring how the tobacco industry adapts to an increasingly restrictive physical marketing environment through the use of URLs on packs and other innovations could be used to inform increased packaging restrictions and plain packaging to reduce the industry's ability to link tobacco users with additional protobacco messaging. Future marketing restrictions should explicitly include the online environment, and surveillance efforts should monitor industry websites to ensure proper enforcement of existing marketing restrictions.

Limitations and Strengths

The codebook for this project assessed a finite number of topics. Additionally, we did not code every webpage, leaving some content undocumented. Given the massive amount of content on websites we would have obtained if we had assessed every possible web component (including both internal and external links), we had to establish a limit on what was possible to code.

A total of 15 websites required translation; while translators were fluent in the translated language and in English, it is possible that there were some inaccuracies. To try to account for any confusion about translation, translators were present either in person or via Skype for clarification. Tobacco packs where the websites were found were purchased in 2013, and at the time of analysis (2016-2017) some websites were inactive, reducing the number of websites in our sample.

We were able to include coding only of corporate websites, nontobacco websites, and a limited number of brand and social media websites due to the terms and conditions regarding who was allowed to register and access website content. This limitation prevented us from deeply exploring brand and social media websites and allowed only a cursory glimpse of the content on these types of websites. There is a need for the research community to recognize and address such restrictions and establish norms for conducting critical public health surveillance research in online spaces.

To our knowledge, this is the most extensive analysis of an international sample of corporate tobacco websites. The sampling frame is unique in that URLs were from cigarette packs purchased in 14 LMICs, therefore covering a broader range of available websites, including state-owned and regional and national company websites in addition to country-specific multinational websites. Research on marketing, specifically internet marketing, in LMICs is rare.

Conclusions

The very existence of tobacco websites, whether corporate, brand, or social media, challenges existing laws concerning tobacco marketing in the countries where packs with URLs were purchased. According to legal analysis from the Tobacco Control Laws website [67] (see Multimedia Appendix 1), at least 7 of the 13 countries where packs were purchased currently ban online tobacco advertising. Bans in China and Thailand were passed after packs were purchased in 2013, but even these 2 later bans were in place before we conducted our content analysis and should have resulted in these websites being removed. An additional 3 countries have legislation that could be interpreted as banning online advertising, and 2 have some restrictions on online advertising. Brand and social media tobacco website URLs appeared on packs purchased in Russia and Vietnam, where online marketing is banned. It's unclear whether corporate websites, which clearly include marketing appeals and engagement strategies, as well as the promotion of CSR programs, are included in these bans. What is clear is that tobacco companies are taking advantage of this lack of clarity in legislation and lack of enforcement to use the internet to further disseminate both corporate and brand marketing that serves to improve their corporate image and keep people using tobacco products. It is essential that marketing regulations become more comprehensive and ban all protobacco communication, a policy that is in line with articles 5.3 and 13 of the FCTC [57,58]. Plain packaging and other packaging restrictions would also limit the industry's ability for mass dissemination of their websites and should also be considered. These policy recommendations are especially important for LMICs, where the tobacco industry has concentrated their marketing efforts to take advantage of the looser regulatory environment [37,68,69]. Efforts to monitor tobacco company marketing should consider including online marketing, as well as social media activities, in light of their growing ubiquity.

Acknowledgments

This work was supported with funding by Bloomberg Philanthropies' Bloomberg Initiative to Reduce Tobacco Use. CVW was supported by T32 CA009314 during manuscript development.

CVW was responsible for study design, coding, reliability calculations, analysis, and manuscript preparation. AYH assisted with coding and manuscript preparation. KS and JEC reviewed the study design and assisted in manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online marketing regulations in 13 low- and middle-income countries where URLs on packs were found.
[PDF File (Adobe PDF File), 78 KB - jmir_v22i6e15160_app1.pdf]

Multimedia Appendix 2
Codebook for quantitative coding of corporate and brand websites.
[DOCX File , 17 KB - jmir_v22i6e15160_app2.docx]

Multimedia Appendix 3
List of active URLs and associated pack information.
[DOCX File , 24 KB - jmir_v22i6e15160_app3.docx]

References

1. Lovato C, Watts A, Stead LF. Impact of tobacco advertising and promotion on increasing adolescent smoking behaviours. Cochrane Database Syst Rev 2011 Oct 05(10):CD003439. [doi: 10.1002/14651858.CD003439.pub2] [Medline: 21975739]
2. Madkour AS, Ledford EC, Andersen L, Johnson CC. Tobacco advertising/promotions and adolescents' smoking risk in Northern Africa. Tob Control 2014 May;23(3):244-252. [doi: 10.1136/tobaccocontrol-2012-050593] [Medline: 23303288]
3. Lee J, Johnson C, Rice J, Warren CW, Chen T. Smoking beliefs and behavior among youth in South Korea, Taiwan, and Thailand. Int J Behav Med 2013 Sep;20(3):319-326. [doi: 10.1007/s12529-012-9236-3] [Medline: 22592594]
4. Hanewinkel R, Isensee B, Sargent JD, Morgenstern M. Cigarette advertising and teen smoking initiation. Pediatrics 2011 Feb;127(2):e271-e278. [doi: 10.1542/peds.2010-2934] [Medline: 21242217]
5. Siahpush M, Shaikh RA, Smith D, Hyland A, Cummings KM, Kessler AS, et al. The association of exposure to point-of-sale tobacco marketing with quit attempt and quit success: results from a prospective study of smokers in the United States. Int J Environ Res Public Health 2016 Feb 06;13(2):203 [FREE Full text] [doi: 10.3390/ijerph13020203] [Medline: 26861379]
6. Germain D, McCarthy M, Wakefield M. Smoker sensitivity to retail tobacco displays and quitting: a cohort study. Addiction 2010 Jan;105(1):159-163. [doi: 10.1111/j.1360-0443.2009.02714.x] [Medline: 19804457]
7. National Cancer Institute. The Role of the Media in Promoting and Reducing Tobacco Use. NCI Tobacco Control Monograph Series No. 19. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health; 2008. URL: https://cancercontrol.cancer.gov/brp/tcrb/monographs/19/m19_complete.pdf [accessed 2020-05-11]
8. Henriksen L, Schleicher NC, Feighery EC, Fortmann SP. A longitudinal study of exposure to retail cigarette advertising and smoking initiation. Pediatrics 2010 Aug;126(2):232-238 [FREE Full text] [doi: 10.1542/peds.2009-3021] [Medline: 20643725]
9. Paynter J, Edwards R. The impact of tobacco promotion at the point of sale: a systematic review. Nicotine Tob Res 2009 Jan;11(1):25-35. [doi: 10.1093/ntr/ntn002] [Medline: 19246438]
10. Robertson L, McGee R, Marsh L, Hoek J. A systematic review on the impact of point-of-sale tobacco promotion on smoking. Nicotine Tob Res 2015 Jan;17(1):2-17 [FREE Full text] [doi: 10.1093/ntr/ntu168] [Medline: 25173775]
11. Federal Trade Commission Cigarette Report for 2017. Washington, DC: Federal Trade Commission; 2019. URL: https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-cigarette-report-2017-federal-trade-commission-smokeless-tobacco-report/ftc_cigarette_report_2017.pdf [accessed 2020-05-11]
12. Bach L. Tobacco Product Marketing on the Internet. Washington DC: Campaign for Tobacco-Free Kids; 2019 Mar 08. URL: https://web.archive.org/web/20191224183402/https://www.tobaccofreekids.org/assets/factsheets/0081.pdf [accessed 2019-04-01]
13. Cruz TB. Monitoring the tobacco use epidemic IV. The vector: tobacco industry data sources and recommendations for research and evaluation. Prev Med 2009 Jan;48(1 Suppl):S24-S34. [doi: 10.1016/j.ypmed.2008.10.002] [Medline: 18976685]
14. McDaniel PA, Malone RE. The role of corporate credibility in legitimizing disease promotion. Am J Public Health 2009 Mar;99(3):452-461 [FREE Full text] [doi: 10.2105/AJPH.2008.138115] [Medline: 19106419]
15. McDaniel PA, Cadman B, Malone RE. African media coverage of tobacco industry corporate social responsibility initiatives. Glob Public Health 2018 Feb 07;13(2):129-143 [FREE Full text] [doi: 10.1080/17441692.2016.1149203] [Medline: 26947737]
16. Peeters S, Gilmore AB. Understanding the emergence of the tobacco industry's use of the term tobacco harm reduction in order to inform public health policy. Tob Control 2015 Mar;24(2):182-189 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051502] [Medline: 24457543]

17. Fooks GJ, Gilmore AB, Smith KE, Collin J, Holden C, Lee K. Corporate social responsibility and access to policy élites: an analysis of tobacco industry documents. PLoS Med 2011 Aug;8(8):e1001076 [FREE Full text] [doi: 10.1371/journal.pmed.1001076] [Medline: 21886485]
18. Tesler LE, Malone RE. Corporate philanthropy, lobbying, and public health policy. Am J Public Health 2008 Dec;98(12):2123-2133. [doi: 10.2105/AJPH.2007.128231] [Medline: 18923118]
19. Holcomb JL, Upchurch RS, Okumus F. Corporate social responsibility: what are top hotel companies reporting? Int J Contemp Hosp Manag 2007;19(6):461-475. [doi: 10.1108/09596110710775129]
20. McDaniel PA, Cadman B, Malone RE. Shared vision, shared vulnerability: a content analysis of corporate social responsibility information on tobacco industry websites. Prev Med 2016 Dec;89:337-344 [FREE Full text] [doi: 10.1016/j.ypmed.2016.05.033] [Medline: 27261411]
21. Lee K, Carrillo Botero N, Novotny T. 'Manage and mitigate punitive regulatory measures, enhance the corporate image, influence public policy': industry efforts to shape understanding of tobacco-attributable deforestation. Global Health 2016 Sep 20;12(1):55 [FREE Full text] [doi: 10.1186/s12992-016-0192-6] [Medline: 27650401]
22. Astuti PAS, Assunta M, Freeman B. Raising generation 'A': a case study of millennial tobacco company marketing in Indonesia. Tob Control 2018 Jul;27(e1):e41-e49. [doi: 10.1136/tobaccocontrol-2017-054131] [Medline: 30042229]
23. Wackowski OA, Lewis MJ, Delnevo CD. Qualitative analysis of Camel Snus' website message board--users' product perceptions, insights and online interactions. Tob Control 2011 Mar;20(2):e1. [doi: 10.1136/tc.2010.037911] [Medline: 20943829]
24. Ilakkuvan V, Cantrell J, Vallone D. 'Action. Adventure. Special offers.': How Marlboro engages consumers on its website. Tob Control 2014 Sep;23(5):455-456. [doi: 10.1136/tobaccocontrol-2013-051115] [Medline: 23853251]
25. Freeman B, Chapman S. Gone viral? Heard the buzz? A guide for public health practitioners and researchers on how Web 2.0 can subvert advertising restrictions and spread health information. J Epidemiol Community Health 2008 Sep;62(9):778-782. [doi: 10.1136/jech.2008.073759] [Medline: 18701726]
26. Smith E. Corporate image and public health: an analysis of the Philip Morris, Kraft, and Nestlé websites. J Health Commun 2012;17(5):582-600 [FREE Full text] [doi: 10.1080/10810730.2011.635776] [Medline: 22420639]
27. Smith EA, Malone RE. Philip Morris's health information web site appears responsible but undermines public health. Public Health Nurs 2008;25(6):554-564 [FREE Full text] [doi: 10.1111/j.1525-1446.2008.00743.x] [Medline: 18950420]
28. Freeman B. New media and tobacco control. Tob Control 2012 Mar;21(2):139-144. [doi: 10.1136/tobaccocontrol-2011-050193] [Medline: 22345236]
29. Freeman B, Chapman S. British American Tobacco on Facebook: undermining Article 13 of the global World Health Organization Framework Convention on Tobacco Control. Tob Control 2010 Jun;19(3):e1-e9 [FREE Full text] [doi: 10.1136/tc.2009.032847] [Medline: 20395406]
30. Ribisl KM, Jo C. Tobacco control is losing ground in the Web 2.0 era: invited commentary. Tob Control 2012 Mar;21(2):145-146. [doi: 10.1136/tobaccocontrol-2011-050360] [Medline: 22345237]
31. McDaniel PA, Malone RE. British American Tobacco's partnership with Earthwatch Europe and its implications for public health. Glob Public Health 2012;7(1):14-28 [FREE Full text] [doi: 10.1080/17441692.2010.549832] [Medline: 21347934]
32. Friedman LC. Philip Morris's website and television commercials use new language to mislead the public into believing it has changed its stance on smoking and disease. Tob Control 2007 Dec;16(6):e9 [FREE Full text] [doi: 10.1136/tc.2007.024026] [Medline: 18048599]
33. Unger JB, Bartsch L. Exposure to tobacco websites: associations with cigarette and e-cigarette use and susceptibility among adolescents. Addict Behav 2018 Mar;78:120-123 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.012] [Medline: 29154150]
34. O'Brien EK, Navarro MA, Hoffman L. Mobile website characteristics of leading tobacco product brands: cigarettes, smokeless tobacco, e-cigarettes, hookah and cigars. Tob Control 2018 Aug 30;8(5). [doi: 10.1136/tobaccocontrol-2018-054549] [Medline: 30166427]
35. Brock B, Carlson SC, Moilanen M, Schillo BA. Reaching consumers: how the tobacco industry uses email marketing. Prev Med Rep 2016 Dec;4:103-106 [FREE Full text] [doi: 10.1016/j.pmedr.2016.05.020] [Medline: 27413669]
36. Timberlake DS, Bruckner TA, Ngo V, Nikitin D. The reach and impact of direct marketing via brand websites of moist snuff. Tob Regul Sci 2016 Apr;2(2):153-165 [FREE Full text] [doi: 10.18001/TRS.2.2.6] [Medline: 27517061]
37. Gilmore AB, Fooks G, Drope J, Bialous SA, Jackson RR. Exposing and addressing tobacco industry conduct in low-income and middle-income countries. Lancet 2015 Mar 14;385(9972):1029-1043 [FREE Full text] [doi: 10.1016/S0140-6736(15)60312-9] [Medline: 25784350]
38. GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet 2017 May 13;389(10082):1885-1906 [FREE Full text] [doi: 10.1016/S0140-6736(17)30819-X] [Medline: 28390697]
39. Lewis MJ, Yulis SG, Delnevo C, Hrywna M. Tobacco industry direct marketing after the Master Settlement Agreement. Health Promot Pract 2004 Jul;5(3 Suppl):75S-83S. [doi: 10.1177/1524839904264596] [Medline: 15231100]
40. Freeman B, Chapman S. Open source marketing: Camel cigarette brand marketing in the “Web 2.0” world. Tob Control 2009 Jun;18(3):212-217. [doi: 10.1136/tc.2008.027375] [Medline: 19211615]

41. Smith K, Washington C, Brown J, Vadnais A, Kroart L, Ferguson J, et al. The Tobacco Pack Surveillance System: a protocol for assessing health warning compliance, design features, and appeals of tobacco packs sold in low- and middle-income countries. JMIR Public Health Surveill 2015;1(2):e8 [FREE Full text] [doi: 10.2196/publichealth.4616] [Medline: 27227142]
42. National Cancer Institute. NCI Dictionary of Cancer Terms: kretek. Rockville, MD: National Cancer Institute, National Institutes of Health, US Department of Health and Human Services; 2019. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/kretek [accessed 2020-05-11]
43. National Cancer Institute. NCI Dictionary of Cancer Terms: bidi. Rockville, MD: National Cancer Institute, National Institutes of Health, US Department of Health and Human Services; 2019. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/bidi [accessed 2020-05-11]
44. Institute for Global Tobacco Control. TPackSS: Tobacco Pack Surveillance System. Baltimore, MD: Johns Hopkins University; 2020. URL: https://globaltobaccocontrol.org/tpackss/ [accessed 2020-05-12]
45. Johns Hopkins Bloomberg School of Public Health, Institute for Global Tobacco Control. Online marketing of electronic cigarettes in Russia: fact sheet. Baltimore, MD: Institute for Global Tobacco Control; 2016. URL: http://globaltobaccocontrol.org/sites/default/files/FS-russia-ecigaretteweb.pdf [accessed 2020-05-11]
46. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
47. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993 May;46(5):423-429. [Medline: 8501467]
48. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. J Clin Epidemiol 1996 Apr;49(4):431-434. [doi: 10.1016/0895-4356(95)00571-4] [Medline: 8621993]
49. Lee K, Eckhardt J. The globalisation strategies of five Asian tobacco companies: an analytical framework. Glob Public Health 2017 Mar;12(3):269-280 [FREE Full text] [doi: 10.1080/17441692.2016.1251604] [Medline: 27884083]
50. Djarum. 2018. URL: https://www.djarum.com/brands/overview/ [accessed 2019-05-02] [WebCite Cache ID 783gvB8OT]
51. British American Tobacco. Sustainability. London, UK: British American Tobacco plc URL: http://www.bat.com/group/sites/uk__9d9kcy.nsf/vwPagesWebLive/DO52AD7G [accessed 2020-05-11]
52. Balbach ED, Smith EA, Malone RE. How the health belief model helps the tobacco industry: individuals, choice, and "information". Tob Control 2006 Dec;15 Suppl 4:iv37-iv43 [FREE Full text] [doi: 10.1136/tc.2005.012997] [Medline: 17130622]
53. Keller KL. Strategic Brand Management: Building, Measuring, and Managing Brand Equity. 4th edition. Boston, MA: Pearson; 2013.
54. Dorfman L, Cheyne A, Friedman LC, Wadud A, Gottlieb M. Soda and tobacco industry corporate social responsibility campaigns: how do they compare? PLoS Med 2012;9(6):e1001241 [FREE Full text] [doi: 10.1371/journal.pmed.1001241] [Medline: 22723745]
55. Hirschhorn N. Corporate social responsibility and the tobacco industry: hope or hype? Tob Control 2004 Dec;13(4):447-453 [FREE Full text] [doi: 10.1136/tc.2003.006676] [Medline: 15564636]
56. Warner KE. What's a cigarette company to do? Am J Public Health 2002 Jun;92(6):897-900. [doi: 10.2105/ajph.92.6.897] [Medline: 12036774]
57. World Health Organization. Guidelines for implementation of Article 5.3 of the WHO Framework Convention on Tobacco Control. Geneva, Switzerland: WHO; 2008. URL: http://www.who.int/fctc/guidelines/article_5_3.pdf [accessed 2020-05-11]
58. World Health Organization. Guidelines for implementation of Article 13 of the WHO Framework Convention on Tobacco Control (tobacco advertising, promotion, and sponsorship). Geneva, Switzerland: WHO; 2008. URL: http://www.who.int/fctc/guidelines/article_13.pdf?ua=1 [accessed 2020-05-11]
59. Wakefield M, Terry-McElrath Y, Emery S, Saffer H, Chaloupka FJ, Szczypka G, et al. Effect of televised, tobacco company-funded smoking prevention advertising on youth smoking-related beliefs, intentions, and behavior. Am J Public Health 2006 Dec;96(12):2154-2160. [doi: 10.2105/AJPH.2005.083352] [Medline: 17077405]
60. WHO Framework Convention on Tobacco Control. Guidelines for Implementation Article 5.3; Article 8; Articles 9 and 10; Article 11; Article 12; Article 13; Article 14, 2013 edition. Geneva, Switzerland: World Health Organization; 2013. URL: https://apps.who.int/iris/bitstream/handle/10665/80510/9789241505185_eng.pdf?sequence=1&isAllowed=y [accessed 2020-05-11]
61. Escobedo P, Cruz TB, Tsai K, Allem JP, Soto DW, Kirkpatrick MG, et al. Monitoring tobacco brand websites to understand marketing strategies aimed at tobacco product users and potential users. Nicotine Tob Res 2018 Sep 25;20(11):1393-1400. [doi: 10.1093/ntr/ntx200] [Medline: 29059386]
62. DiFranza JR, Wellman RJ, Sargent JD, Weitzman M, Hipple BJ, Winickoff JP, Tobacco Consortium, Center for Child Health Research of the American Academy of Pediatrics. Tobacco promotion and the initiation of tobacco use: assessing the evidence for causality. Pediatrics 2006 Jun;117(6):e1237-e1248. [doi: 10.1542/peds.2005-1817] [Medline: 16740823]
63. Pierce JP, Choi WS, Gilpin EA, Farkas AJ, Berry CC. Tobacco industry promotion of cigarettes and adolescent smoking. JAMA 1998 Feb 18;279(7):511-515. [Medline: 9480360]
64. Biener L, Siegel M. Tobacco marketing and adolescent smoking: more support for a causal inference. Am J Public Health 2000 Mar;90(3):407-411. [doi: 10.2105/ajph.90.3.407] [Medline: 10705860]

65. Choi K, Forster JL. Frequency and characteristics associated with exposure to tobacco direct mail marketing and its prospective effect on smoking behaviors among young adults from the US Midwest. Am J Public Health 2014 Nov;104(11):2179-2183 [FREE Full text] [doi: 10.2105/AJPH.2014.302123] [Medline: 25211739]
66. Hafez N, Ling PM. Finding the Kool Mixx: how Brown & Williamson used music marketing to sell cigarettes. Tob Control 2006 Oct;15(5):359-366 [FREE Full text] [doi: 10.1136/tc.2005.014258] [Medline: 16998169]
67. Campaign for Tobacco-Free Kids. Legislation. Washington, DC: Campaign for Tobacco-Free Kids; 2019. URL: https://www.tobaccocontrollaws.org/legislation [accessed 2020-05-11]
68. Doku D. The tobacco industry tactics-a challenge for tobacco control in low and middle income countries. Afr Health Sci 2010 Jun;10(2):201-203 [FREE Full text] [Medline: 21326977]
69. Lee S, Ling PM, Glantz SA. The vector of the tobacco epidemic: tobacco industry practices in low and middle-income countries. Cancer Causes Control 2012 Mar;23 Suppl 1:117-129 [FREE Full text] [doi: 10.1007/s10552-012-9914-0] [Medline: 22370696]

Abbreviations

CSR: corporate social responsibility
FCTC: Framework Convention on Tobacco Control
LMICs: low- and middle-income countries
TPackSS: Tobacco Pack Surveillance System
WHO: World Health Organization

Edited by G Eysenbach; submitted 24.06.19; peer-reviewed by C Watts, J Villegas; comments to author 03.10.19; revised version received 22.01.20; accepted 27.02.20; published 09.06.20.

Please cite as:
Weiger CV, Smith K, Hong AY, Cohen JE
Cigarette Packs With URLs Leading to Tobacco Company Websites: Content Analysis
J Med Internet Res 2020;22(6):e15160
URL: https://www.jmir.org/2020/6/e15160
doi:10.2196/15160
PMID:32459649

Original Paper

Privacy-Preserving Record Linkage of Deidentified Records Within a Public Health Surveillance System: Evaluation Study

Long Nguyen[1], MCom; Mark Stoové[1,2], PhD; Douglas Boyle[3], PhD; Denton Callander[4,5], PhD; Hamish McManus[4], PhD; Jason Asselin[1], MEpi; Rebecca Guy[4], PhD; Basil Donovan[4,6], MD; Margaret Hellard[1,2,7,8], MD, PhD; Carol El-Hayek[1,2], MEpi

[1]Burnet Institute, Melbourne, Australia
[2]School of Population Health and Preventive Medicine, Monash University, Melbourne, Australia
[3]Department of General Practice, HaBIC Research Technology Unit, Melbourne Medical School, University of Melbourne, Melbourne, Australia
[4]Kirby Institute, University of New South Wales, Sydney, Australia
[5]Mailman School of Public Health, Columbia University, New York, NY, United States
[6]Sydney Sexual Health Centre, Sydney, Australia
[7]Department of Infectious Diseases, The Alfred Hospital, Melbourne, Australia
[8]Doherty Institute and School of Population and Global Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Long Nguyen, MCom
Burnet Institute
85 Commercial Rd
Melbourne, 3004
Australia
Phone: 61 468689589
Email: long.nguyen@burnet.edu.au

Abstract

Background: The Australian Collaboration for Coordinated Enhanced Sentinel Surveillance (ACCESS) was established to monitor national testing and test outcomes for blood-borne viruses (BBVs) and sexually transmissible infections (STIs) in key populations. ACCESS extracts deidentified data from sentinel health services that include general practice, sexual health, and infectious disease clinics, as well as public and private laboratories that conduct a large volume of BBV/STI testing. An important attribute of ACCESS is the ability to accurately link individual-level records within and between the participating sites, as this enables the system to produce reliable epidemiological measures.

Objective: The aim of this study was to evaluate the use of GRHANITE software in ACCESS to extract and link deidentified data from participating clinics and laboratories. GRHANITE generates irreversible hashed linkage keys based on patient-identifying data captured in the patient electronic medical records (EMRs) at the site. The algorithms to produce the data linkage keys use probabilistic linkage principles to account for variability and completeness of the underlying patient identifiers, producing up to four linkage key types per EMR. Errors in the linkage process can arise from imperfect or missing identifiers, impacting the system's integrity. Therefore, it is important to evaluate the quality of the linkages created and evaluate the outcome of the linkage for ongoing public health surveillance.

Methods: Although ACCESS data are deidentified, we created two gold-standard datasets where the true match status could be confirmed in order to compare against record linkage results arising from different approaches of the GRHANITE Linkage Tool. We reported sensitivity, specificity, and positive and negative predictive values where possible and estimated specificity by comparing a history of HIV and hepatitis C antibody results for linked EMRs.

Results: Sensitivity ranged from 96% to 100%, and specificity was 100% when applying the GRHANITE Linkage Tool to a small gold-standard dataset of 3700 clinical medical records. Medical records in this dataset contained a very high level of data completeness by having the name, date of birth, post code, and Medicare number available for use in record linkage. In a larger gold-standard dataset containing 86,538 medical records across clinics and pathology services, with a lower level of data completeness, sensitivity ranged from 94% to 95% and estimated specificity ranged from 91% to 99% in 4 of the 6 different record linkage approaches.

Conclusions: This study's findings suggest that the GRHANITE Linkage Tool can be used to link deidentified patient records accurately and can be confidently used for public health surveillance in systems such as ACCESS.

(J Med Internet Res 2020;22(6):e16757) doi:10.2196/16757

KEYWORDS

medical record linkage; public health surveillance; sentinel surveillance; sensitivity and specificity; data linkage; confidentiality; evaluation studies as a topic

Introduction

Background

The Australian Collaboration for Coordinated Enhanced Sentinel Surveillance (ACCESS) of blood-borne viruses (BBVs) and sexually transmissible infections (STIs) monitors diagnostic testing and other episodes of care for priority BBVs and STIs [1,2]. ACCESS extracts deidentified patient data from a network of laboratories and clinics, including those that manage high caseloads of patients from populations with heightened risk for BBVs and STIs, including gay, bisexual, and other men who have sex with men and people who inject drugs. The main objective of ACCESS is to measure and report key indicators such as disease incidence and prevalence (measured as proportion positive) to monitor and support Australia's efforts to reduce the transmission of morbidity and mortality caused by BBV and STI [3-5].

A key challenge for ACCESS (and similar sentinel surveillance systems) is that patient outcomes can be inaccurately measured if individuals attend multiple health services, leading to potential reporting bias. For example, markers of testing frequency, an important indicator for BBV/STI prevention and management [3-5], may be underestimated if individuals test at multiple services. Accurate linkage of individuals within and between services in ACCESS provides more accurate measures of (1) the key indicators relating to Australia's BBV and STI National Strategies and (2) the effect of interventions aimed at reducing the impact of BBVs and STIs.

The linkage of deidentified ACCESS records across sites relies on specialized health data extraction software GRHANITE, which is installed at participating clinics and laboratories. GRHANITE interfaces with patient databases, securely extracting line-listed consultation, demographic, BBV and STI clinical and pathology data [6]. Before data are transferred from ACCESS sites, GRHANITE creates a unique record ID to identify an electronic medical record (EMR) and uses patient-identifying information to generate irreversible hash-coded linkage keys associated with that record. The record ID and linkage keys, rather than the personal identifiers, are transferred with the patients' clinical and pathology data to a secure data bank, preserving patient privacy [7]. A link ID is then generated [8] when the same patient is linked across different EMRs by matching linkage keys using a companion software called the GRHANITE Linkage Tool [9].

Objectives

The GRHANITE Linkage Tool has been validated to perform large-scale population-level record linkage [10] to achieve similar sensitivity and specificity data linkage profiles as per traditional person-identifiable data linkage mechanisms [9]. Given that there is variation in the available person-identifiable data at clinical and laboratory sites in ACCESS, the focus of this paper is to assess the quality of linkage results by applying different approaches to using the GRHANITE Linkage Tool in ACCESS. To evaluate the GRHANITE Linkage Tool for ongoing public health surveillance, we measured the outcomes of record linkage using the tool against gold-standard linked datasets derived from ACCESS data.

Methods

Australian Collaboration for Coordinated Enhanced Sentinel Surveillance Data Extraction and Linkage via GRHANITE

Electronic Medical Records

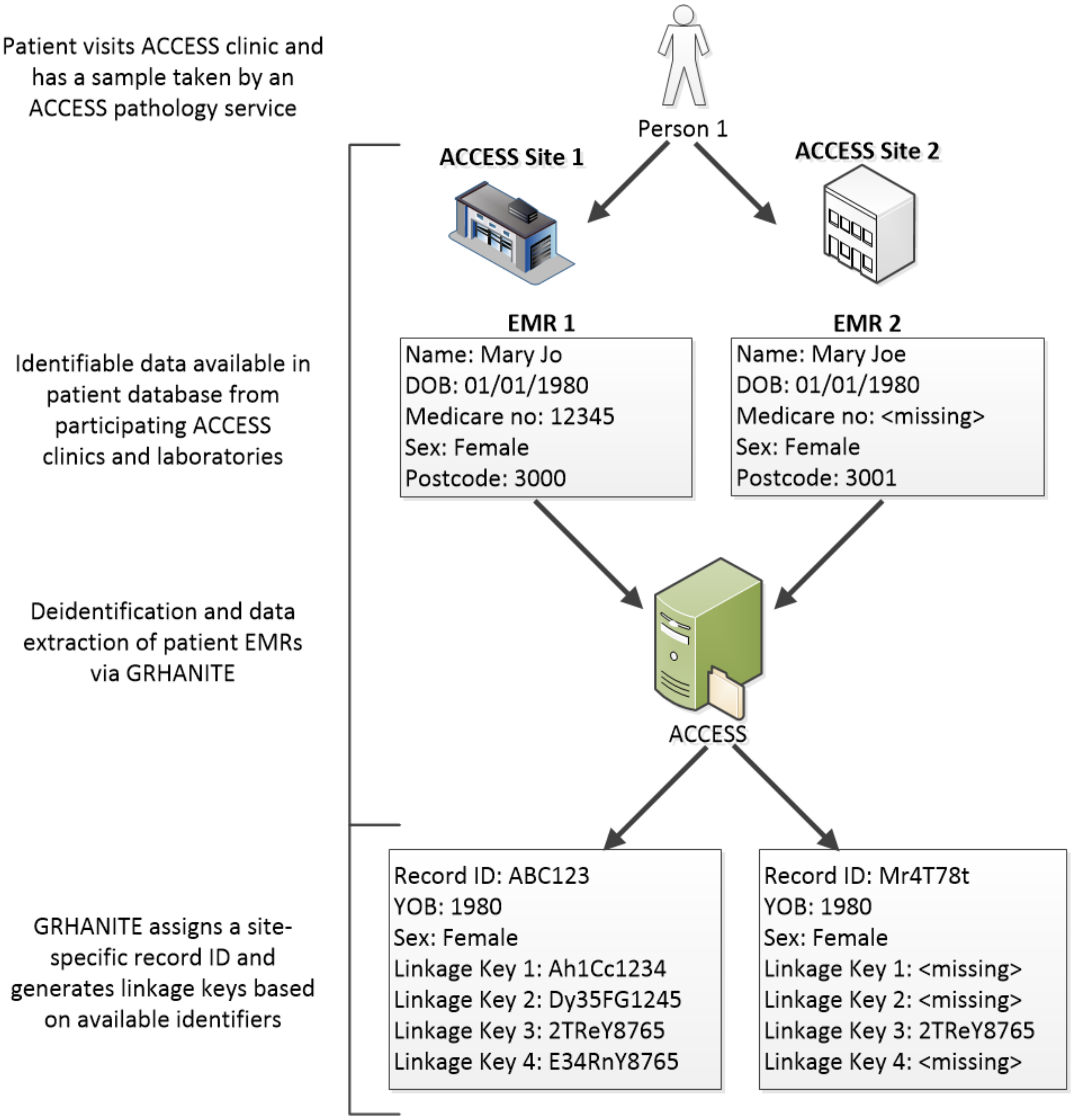
Typically, when a patient first attends a medical facility, an EMR is created in the facility's patient database, containing the patient's identifying information, including the name, date of birth, contact details, and Medicare number (an Australian government–issued health care card number used for Medicare billing). Most clinics will also have recorded other demographic information, such as preferred language, country of birth, and indigenous background in the EMR.

Every individual's EMR has a unique medical record number generated by the patient database, linking all of a patient's consultations, tests, and prescription records. Multiple EMRs may be created for one patient at the same facility if the patient's details change and are not updated, leading to the creation of a new EMR; if the patient uses an alias; or if the patient attends a clinic that allows anonymous or free testing.

Data Extraction in the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance

Data were extracted from participating ACCESS clinical sites that included an EMR for every patient available in their databases at the time of extraction. GRHANITE generated a new unique record ID and up to four irreversible hash-coded linkage keys for each EMR. Personal identifying information (eg, name, date of birth, Medicare number) in the patient's EMR was passed through advanced encryption to generate both record ID and linkage keys [7]. The record ID and linkage keys were extracted by GRHANITE alongside the patient demographics, consultation, test request, pathology, and prescription information related to BBV and STI care, without the identifying information. Data extraction was similar for laboratories; however, only BBV and STI test records related to diagnosis and care and a limited set of demographic variables were available for extraction (Figure 1).

Figure 1. Data extraction in the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance: using GRHANITE to deidentify electronic medical records and create linkage keys.



EMR 1
Name: Mary Jo
DOB: 01/01/1980
Medicare no: 12345
Sex: Female
Postcode: 3000

EMR 2
Name: Mary Joe
DOB: 01/01/1980
Medicare no: <missing>
Sex: Female
Postcode: 3001

Record ID: ABC123
YOB: 1980
Sex: Female
Linkage Key 1: Ah1Cc1234
Linkage Key 2: Dy35FG1245
Linkage Key 3: 2TReY8765
Linkage Key 4: E34RnY8765

Record ID: Mr4T78t
YOB: 1980
Sex: Female
Linkage Key 1: <missing>
Linkage Key 2: <missing>
Linkage Key 3: 2TReY8765
Linkage Key 4: <missing>

Record Linkage in the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance

The data components used by GRHANITE to create the linkage keys include the following patient identifiers: 5 digits of the Medicare number, date of birth, sex, first name, last name, and residential postcode. However, not all EMRs have the same set of patient identifiers recorded in the same way. For example, a patient name may be recorded as *William* in one clinic with a full date of birth and *Bill* in another clinic with only a year of birth recorded. GRHANITE utilizes data preprocessing to remove unwanted characters and words and to resolve nicknames utilizing an Australian national nickname list. Phonetic encoding (double metaphone) is then employed, which permits fuzzy matching based on misspellings of the surname and forename. Transposition of day and month of birth is also supported. After preprocessing, identifiers are combined and then encrypted utilizing secret seeding keys and cryptographic hashing to generate the GRHANITE privacy-preserving cryptographic hashed linkage keys [7,9].

GRHANITE creates up to four linkage keys for each EMR, using combinations of identifying information that is recorded at each site (Textbox 1) [11]. For example, if the Medicare number was not recorded for a patient, then linkage keys that require 5 Medicare digits (Textbox 1: linkage key types 1, 2, and 4) could not be created, resulting in EMRs extracted via GRHANITE having only one linkage key (Textbox 1: linkage key 3 does not require the Medicare number; Figure 1).

Textbox 1. Types of linkage keys generated by GRHANITE.

Linkage key and components of base identifying information:

- Type 1: 5 Medicare digits; date of birth; and sex
- Type 2: 5 Medicare digits; postcode; first three characters of first name; and year of birth
- Type 3: Last name and first name (either order permitted) and fuzzy matching used; date of birth with day/month (transpositions permitted)
- Type 4: Last name and first name (either order permitted) and fuzzy matching used; 5 Medicare digits

Applying the GRHANITE Linkage Tool

There are three steps in the record linkage process in ACCESS when applying the linkage tool. The first step finds pairs of EMRs based on at least one linkage key matching and records the linkage key type/s used to match each record pair. The second step examines the strength of the link using other available data within the matched pair of records to accept or reject linked records as described in Table 1. The third step assigns an identifier (a link ID) to the accepted matched pairs to label all matched records as belonging to the same individual (Figure 2).

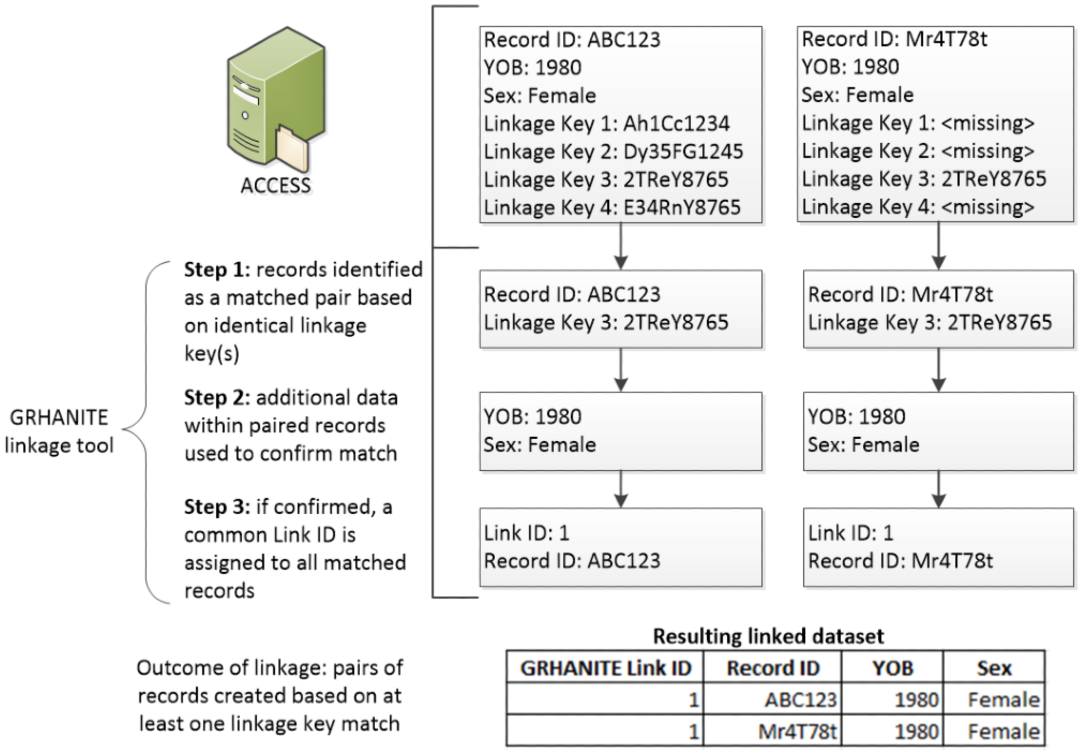
Table 1. GRHANITE Linkage Tool approaches to accepting matches.

Linkage approach	Description
Accept all	Accept all record links as determined by the linkage tool
Year of birth match	Accept only record links if year of birth matches
Sex match	Accept only record links if sex matches
Year of birth and sex match	Accept only record links if year of birth and sex match
Two or more linkage keys[a]	Accept record links only if matched on two or more linkage key types
Linkage key type 3 plus sex match[b]	Accept only record links that match on linkage key type 3 and match on sex

[a] Given that 3 out of the 4 linkage key types are generated using the Medicare number, this approach requires the Medicare number to be present in the EMR (Textbox 1).

[b] This approach only relies on linkage key type 3, which does not require the Medicare number to be present in the EMR.

Figure 2. Record linkage in the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance: using the GRHANITE Linkage Tool to identify and accept matches.



Record ID: ABC123
YOB: 1980
Sex: Female
Linkage Key 1: Ah1Cc1234
Linkage Key 2: Dy35FG1245
Linkage Key 3: 2TReY8765
Linkage Key 4: E34RnY8765

Record ID: Mr4T78t
YOB: 1980
Sex: Female
Linkage Key 1: <missing>
Linkage Key 2: <missing>
Linkage Key 3: 2TReY8765
Linkage Key 4: <missing>

Step 1: records identified as a matched pair based on identical linkage key(s)

Record ID: ABC123
Linkage Key 3: 2TReY8765

Record ID: Mr4T78t
Linkage Key 3: 2TReY8765

Step 2: additional data within paired records used to confirm match

YOB: 1980
Sex: Female

YOB: 1980
Sex: Female

Step 3: if confirmed, a common Link ID is assigned to all matched records

Link ID: 1
Record ID: ABC123

Link ID: 1
Record ID: Mr4T78t

Outcome of linkage: pairs of records created based on at least one linkage key match

Resulting linked dataset

GRHANITE Link ID	Record ID	YOB	Sex
1	ABC123	1980	Female
1	Mr4T78t	1980	Female

Evaluating the Record Linkage

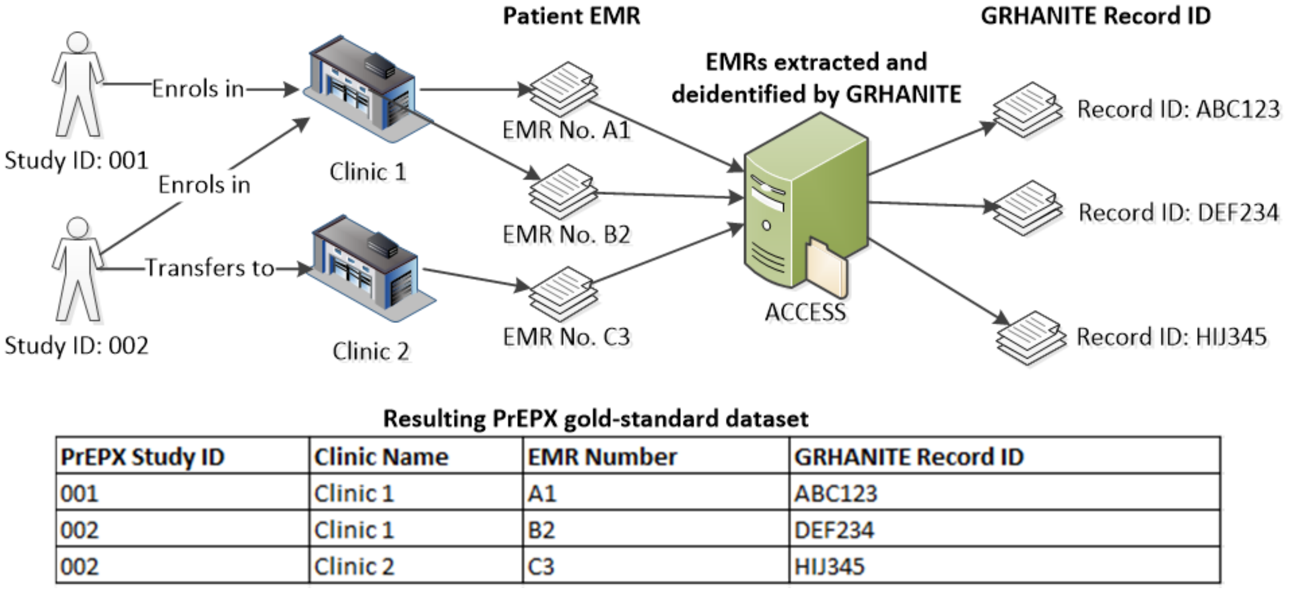
Creating the Gold-Standard Datasets

To evaluate the record linkage in ACCESS, we generated two gold-standard datasets, using a deterministic record linkage method, where the true match status could be identified [12]. To assess the outcomes of the six linkage approaches described in Table 1, using the GRHANITE Linkage Tool for matching the records in the gold-standard datasets, we measured the sensitivity, specificity, and positive and negative predictive values, where possible.

The PrEPX Gold-Standard Dataset

PrEPX is a population-level intervention study in Victoria in which HIV pre-exposure prophylaxis was made available to eligible individuals, and the study used ACCESS data to monitor participants' BBV and STI testing [13]. Eight clinical sites and one hospital clinic participating in ACCESS had PrEPX participants enrolled between July 2016 and March 2018. At enrollment, a PrEPX study ID was sequentially assigned and recorded alongside each participant's enrollment-clinic EMR number in a study database. Following enrollment in PrEPX, if a participant attended a different clinic within the network during the study period, the EMR number from the new clinic was also recorded in the study database. ACCESS had ethics approval to extract the EMR number from the participating clinics for the purpose of matching the records of participants who moved among clinics. To create the gold-standard dataset, EMRs were matched on clinic EMR number, clinic name, and clinic visit date. The gold standard included one record per PrEPX participant who attended only one clinic during the study period and multiple records per PrEPX participant who attended multiple clinics linked by study ID (Figure 3).

Figure 3. Data flow of electronic medical records in PrEPX and deterministic linkage for the gold-standard dataset.



Resulting PrEPX gold-standard dataset

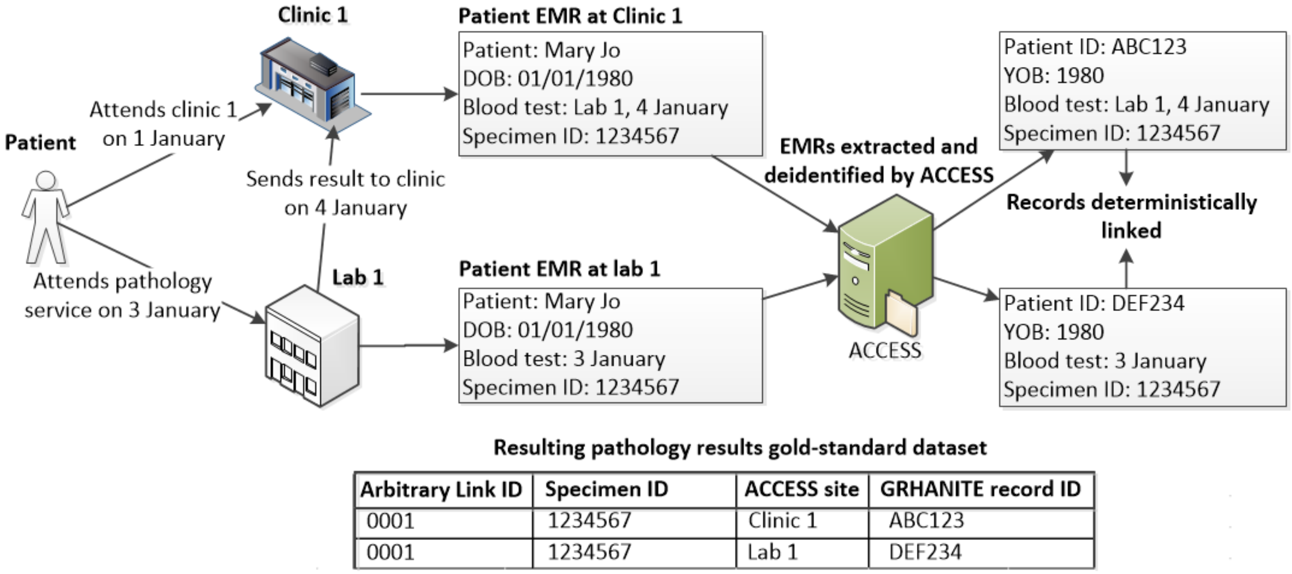
PrEPX Study ID	Clinic Name	EMR Number	GRHANITE Record ID
001	Clinic 1	A1	ABC123
002	Clinic 1	B2	DEF234
002	Clinic 2	C3	HIJ345

The Pathology Results Gold-Standard Dataset

A second and much larger gold-standard dataset was generated from the EMRs extracted from 7 clinics and 4 laboratories participating in ACCESS between January 2009 and April 2018. To be included in this dataset, patients had to have at least one specimen sent from one of the ACCESS clinics to one of the ACCESS laboratories. A unique laboratory specimen ID was assigned to the specimen at the laboratory, and when laboratories returned pathology results to the clinic, this specimen ID was also recorded at the clinic. To create the gold-standard dataset, clinic and laboratory records were matched using the laboratory specimen ID, year of birth, and test date. We allowed for a 7-day difference in test dates, as in medical records, the recorded date can commonly vary for the same specimen. Only matched records were included in the gold-standard dataset and linked using an arbitrarily assigned link identifier (Figure 4).

An EMR in the pathology results gold-standard dataset may match to many other EMRs for several reasons, including the following: individuals may have had multiple specimens sent to multiple laboratories for testing, individuals may have attended different clinics and therefore had the same test result sent from the laboratory to more than one clinic, or individuals may have had multiple EMRs at the laboratory or clinic as a result of outdated or incomplete personal identifiers.

Figure 4. Data flow of pathology results in electronic medical records and deterministic linkage for the gold-standard dataset.



Arbitrary Link ID	Specimen ID	ACCESS site	GRHANITE record ID
0001	1234567	Clinic 1	ABC123
0001	1234567	Lab 1	DEF234

Data Analysis

Sensitivity

The sensitivity was calculated as the number of correctly linked EMRs, as identified using the GRHANITE Linkage Tool, as a percentage of the total number of linked EMRs in the gold standard dataset.

Specificity

In the PrEPX gold-standard dataset, the specificity was calculated as the number of single EMRs correctly identified as unlinked using the GRHANITE Linkage Tool as a percentage of the total number of unlinked EMRs. The positive predictive value (PPV) and negative predictive value were also calculated to provide probabilities of true matches and missed matches.

Given the deidentified nature of the ACCESS data, it was not possible to include unmatched specimen IDs in the pathology results gold-standard dataset because there was no way to confirm whether they belonged to different individuals (correctly unmatched), making it impossible to calculate specificity. Therefore, to evaluate specificity, we assessed the concordance of chronological HIV and hepatitis C test records to identify EMRs that should not have been linked. By identifying the linked EMRs with discordant results, the PPV (the proportion of linked records with concordant antibody results) could be determined. The specificity was then estimated using the PPV and the sensitivity for each linkage approach as summarized in Figure 5.

Figure 5. Estimating specificity when positive predictive value and sensitivity are known.

		Gold-Standard	
		Linked EMR	Unlinked EMR
Observed GRHANITE Linkage Result	Linked EMR	True Positive (TP)	False Positive (FP)
	Unlinked EMR	False Negative (FN)	True Negative (TN)

When PPV is known, the number of false and true positives can be derived:

- FP = (1-PPV) x Observed number of linked EMRs
- TP = Observed number of linked EMRs – FP

When sensitivity is known, the number of false and true negatives can be derived:

- FN = (TP/Sensitivity) - TP
- TN = Observed number of unlinked EMRs – FN

The number of true negative and false positives can then be used to estimate specificity:

- Specificity = [TN/(FP + TN)] × 100

Measuring Incorrect Matches Using Discordant Pathology Results

Following infection, any HIV or hepatitis C antibody test that subsequently occurs should always return a positive result. Using the pathology results gold-standard dataset provided only a small number of HIV and hepatitis C results; therefore, a dataset of linked EMRs was derived using all available EMRs from the same clinic and laboratory sites used to create the gold-standard dataset. Two datasets were created, one that contained any HIV western blot or antibody result and one that contained any hepatitis C antibody result. EMRs containing discordant results before record linkage were excluded from the sample so as not to confuse it with discordance resulting from record linkage. Records within each dataset were then linked using all six approaches (Table 1) of the GRHANITE Linkage Tool. Linked EMRs where there was no history of a positive result were removed from the sample, as a discordant test result can only be determined after an initial positive result. Therefore, in the HIV and hepatitis C datasets, only linked EMRs that contained an antibody result after an initial positive HIV western blot or hepatitis C antibody result were retained for evaluation.

To calculate the PPV, the linked EMRs were then inspected for negative antibody results occurring at least seven days after a positive test result, which were then classified as incorrectly matched. Where most subsequent antibody tests were negative, the initial and any subsequent positive results were considered incorrectly matched records.

Results

Record Linkage Using the PrEPX Gold-Standard Dataset

The PrEPX gold-standard dataset identified 28 joins among 56 EMRs, indicating 28 study participants had attended two different clinical sites during the PrEPX study period. The remaining 3644 EMRs were from participants who only attended a single clinic during the study and therefore did not have any linked records.

Over 99% of EMRs had all four linkage key types present in 8 of the 9 sites, indicating that the patient-identifying information to generate those linkage keys was near fully recorded at the clinics. One site was missing data needed to generate linkage types 1, 2, and 4 (which all require the Medicare number) in 11% (8/76) of their EMRs (Table 2).

In all linkage approaches, except the approach requiring two or more linkage keys, all pairs of EMRs from the 28 individuals who attended two sites were correctly joined (100% sensitivity). With the approach which required two or more linkage keys for matching, one pair was not identified (96% sensitivity). Specificity was 100% using all linkage approaches, without any of the remaining 3644 EMRs in the dataset being falsely linked (Table 3).

Table 2. Percentage of electronic medical records in the PrEPX gold-standard dataset by linkage key type and site.

Site	Number of electronic medical records, N	Percentage of electronic medical records with Linkage Key			
		Type 1, n (%)	Type 2, n (%)	Type 3, n (%)	Type 4, n (%)
Site 1	76	68 (89)	68 (89)	76 (100)	68 (89)
Site 2	853	853 (100.0)	853 (100.0)	853 (100.0)	853 (100.0)
Site 3	1087	1087 (100.00)	1084 (99.72)	1087 (100.00)	1087 (100.00)
Site 4	582	582 (100.0)	582 (100.0)	582 (100.0)	582 (100.0)
Site 5	40	40 (100.0)	40 (100.0)	40 (100.0)	40 (100.0)
Site 6	135	135 (100.0)	135 (100.0)	135 (100.0)	135 (100.0)
Site 7	106	106 (100.0)	103 (99.2)	106 (100.0)	106 (100.0)
Site 8	314	314 (100.0)	314 (100.0)	314 (100.0)	314 (100.0)
Site 9	507	507 (100.0)	507 (100.0)	507 (100.0)	507 (100.0)
Total	3700	3692 (99.78)	3686 (99.62)	3700 (100.00)	3692 (99.78)

Table 3. Evaluation measures derived from using the GRHANITE Linkage Tool on the PrEPX gold-standard dataset.

Linkage approach	Sensitivity (N=56), n (%)	Specificity (N=3644), n (%)	Positive predictive value (N=56), n (%)	Negative predictive value (N=3644), n (%)
Accept all	56 (100)	3644 (100.00)	56 (100)	3644 (100.00)
Year of birth match	56 (100)	3644 (100.00)	56 (100)	3644 (100.00)
Sex match	56 (100)	3644 (100.00)	56 (100)	3644 (100.00)
Year of birth and sex match	56 (100)	3644 (100.00)	56 (100)	3644 (100.00)
Two or more linkage keys	54 (96)	3644 (100.00)	54 (100)[a]	3644 (99.90)[b]
Linkage key type 3 plus sex match	56 (100)	3644 (100.00)	56 (100)	3644 (100.00)

[a] N=54.

[b] N=3646.

Record Linkage Using the Pathology Results Gold-Standard Dataset

Using the GRHANITE Linkage Tool on the pathology results gold-standard dataset created 50,484 linked records among 86,538 EMRs, with a maximum of six EMRs identified as belonging to the same individual.

A total of 99.69% (86,273/86,538) of EMRs contained at least one linkage key type, and all four linkage key types were present in 73.51% (63,610/86,538) of records, suggesting that the completion of patient-identifying information in the patient database was very high overall. However, 21.62% (18,709/86,538) of EMRs had only linkage key type 3 available for matching. One or more of linkage types 1, 2, and 4 (which all require the Medicare number) was missing in 97.42% (7914/8124) of EMRs from one public laboratory, 53.95% (5967/11,060) of EMRs from the sexual health clinic, 48.25% (1403/2908) of EMRs from a private laboratory, and 23.42% (6134/26,186) of EMRs from another public laboratory (Table 4).

For the first 4 linkage approaches, the GRHANITE Linkage Tool correctly linked 94% to 95% of EMRs in the pathology results gold-standard dataset, dropping to 66% (57,330/86,538) where two or more linkage keys are needed to form a match (Table 5). In the final linkage approach, where pairs were only accepted when matched on linkage key type 3 (which does not require the Medicare number) and sex, 89% (76,928/86,538) of records were correctly linked.

Table 4. Percentage of electronic medical records in the pathology gold-standard dataset by linkage key type and site.

Site	Number of electronic medical records, N	Number of electronic medical records with no linkage keys, n (%)	Percentage of electronic medical records with Linkage Key			
			Type 1, n (%)	Type 2, n (%)	Type 3, n (%)	Type 4, n (%)
Clinic 1	3165	0 (0.00)	3083 (97.41)	3077 (97.22)	3165 (100)	3083 (97.41)
Clinic 2	6342	0 (0.00)	6031 (95.10)	6015 (94.84)	6342 (100)	6031 (95.10)
Clinic 3	2514	0 (0.00)	2493 (99.16)	2489 (99.01)	2513 (99.96)	2492 (99.12)
Clinic 4	9679	0 (0.00)	9351 (96.61)	9322 (96.31)	9676 (99.97)	9350 (96.60)
Clinic 5	1369	1 (0.07)	1357 (99.12)	1356 (99.05)	1368 (99.93)	1357 (99.12)
Clinic 6	2489	5 (0.20)	2315 (93.01)	2288 (91.92)	2484 (99.80)	2315 (93.01)
Clinic 7 (sexual health)	11,060	9 (0.08)	5097 (46.08)	5094 (46.06)	11,049 (99.90)	5095 (46.07)
Lab 1 (public)	26,186	241 (0.92)	23,705 (90.53)	20,059 (76.60)	25,465 (97.25)	23,227 (88.70)
Lab 2 (public)	8124	8 (0.10)	215 (2.65)	210 (2.58)	8116 (99.90)	215 (2.65)
Lab 3 (private)	2908	1 (0.03)	1706 (58.67)	1509 (51.89)	2907 (99.97)	1710 (58.80)
Lab 4 (private)	12,702	0 (0.00)	12,205 (96.09)	12,203 (96.07)	12,700 (99.98)	12,203 (96.07)
Total	86,538	265 (0.31)	67,558 (78.07)	63,622 (73.52)	85,785 (99.13)	67,078 (77.51)

Table 5. Evaluation measures derived from using the GRHANITE Linkage Tool on the pathology results gold-standard dataset.

Linkage approach	Gold standard (N=86,538)	HIV results			Hepatitis C results		
	Sensitivity, n (%)	N	Positive predictive value, n (%)	Estimated specificity, (%)	N	Positive predictive value, n (%)	Estimated specificity, (%)
Accept all	82,345 (95.15)	1427	1245 (87.25)	90.52	3908	3866 (98.93)	99.32
Year of birth match	82,212 (95.00)	1412	1234 (87.39)	90.71	3817	3777 (98.95)	99.34
Sex match	81,689 (94.40)	1257	1143 (90.93)	93.20	3810	3775 (99.08)	99.42
Year of birth and sex match	81,560 (94.25)	1263	1152 (91.21)	93.42	3775	3741 (99.10)	99.43
Two or more linkage keys	57,330 (66.25)	257	256 (99.6)	99.74	2809	2795 (99.50)	99.67
Linkage key type 3 plus sex match	76,928 (88.90)	1090	984 (90.28)	92.98	3626	3596 (99.17)	99.49

Estimating Specificity Using Discordant Test Results

In the derived HIV dataset, the number of linked EMRs containing an initial positive Western blot result ranged from 1090 to 1427 with all linkage approaches except when two or more linkage keys are needed. The linkage approach which requires two or more linkage keys to match resulted in 257 linked EMRs. The PPV was between 87% and 91% for the first 4 linkage approaches and estimated specificity ranged from 90% to 93%. When fewer EMRs were linked because of the different linkage approaches, both PPV and specificity improved (Table 5).

In the derived hepatitis C dataset, with the first 4 linkage approaches, in excess of 3700 linked EMRs contained an initial positive hepatitis C antibody result, with a drop to 2809 records when two or more linkage keys are needed. The PPV was greater than 98.9% and an estimated specificity was over 99% for all six linkage approaches (Table 5).

Discussion

Principal Findings

This paper describes a comprehensive evaluation of a system of probabilistic record linkage using a privacy-preserving software tool within a large-scale health surveillance system. The results showed that this software provides a highly reliable and accurate system for linking routinely collected EMRs through the generation of linkage keys reliant on available identifying information. Optimizing the record linkage involves an appropriate balance between the sensitivity (correctly identifying records belonging to the same person) and specificity (ensuring records that belong to different people are not linked) as well as what will best suit the study design objectives and populations under study without impeding the interpretation of study results.

The high performance of the linkage tool when applied to the relatively small PrEPX gold-standard dataset was related to the data completeness for EMRs in the PrEPX trial compared with the completeness of data in the pathology results gold-standard dataset (Tables 2 and 4). Participants in PrEPX were required to have a Medicare number to be enrolled and have three monthly follow-up visits, which allowed multiple opportunities for the staff at clinics to record any missing identifying information [13]. Where the underlying identifiers are robust and duplication is at a minimum, the probability of missed matches is negligible. In addition, with the PrEPX gold-standard dataset, there was 100% specificity for all linkage approaches, indicating that the linkage tool does not falsely link records in a small sample of EMRs where there was unlikely to be individuals with similar identifying details (name, date of birth, and Medicare number).

When the linkage tool was applied to the larger pathology results gold-standard dataset, sensitivity ranged between 89% and 95% where the linkage approach relied on a single linkage key matching. However, with the approach that requires records to link on two or more linkage key types, sensitivity was reduced to 66%. This is attributable to 22% of EMRs only having a single linkage key type available for linkage, which is mostly because of the Medicare number not being available. The inclusion of laboratory records in the pathology results gold-standard dataset may contribute to a lower sensitivity as a result of patient identifier errors such as mislabeling and recording of laboratory samples [14], compared with the completeness of personal identifiers within clinic EMRs. The final linkage approach where pairs of EMRs were only linked when matched on linkage key type 3 (which does not require the Medicare number) and sex, resulted in 89% sensitivity. This approach was included in the analysis to simulate the performance of the linkage tool when the Medicare number is not available. This is important to evaluate in Australia as a significant proportion of participating sites within ACCESS are funded through jurisdictional governments and do not record patient Medicare numbers [15].

Limitations

The main challenge in evaluating the GRHANITE Linkage Tool was the development of gold-standard datasets given the deidentified nature of EMRs in ACCESS. Researchers rarely have access to gold-standard datasets on which to perform linkage validation outside large administrative health data sources, and our gold-standard dataset of 86,538 records was comparable with other published studies [16]. The gold standards required records with enough supplementary information for deterministic matching where we could be certain that matches belonged to the same individual and nonmatches belonged to different individuals. Therefore, to generate the gold-standard datasets, there were a limited number of records we could use to accurately calculate sensitivity and specificity of the linkage tool. Although the pathology results gold-standard dataset contained over 80,000 records, one limitation of the evaluation was the inability to identify the correctly unmatched EMRs, which meant specificity could not be directly measured. However, given ACCESS is focused on the surveillance of BBV and STI, we were able to evaluate specificity within the pathology results dataset by examining the concordance of linked test results for HIV and hepatitis C. As expected, linkage specificity was inversely related to sensitivity. In addition, using discordant antibody results, we assumed that any discordant result was attributable to incorrect record linkage as opposed to an error in laboratory test results. However, given the very high sensitivity and specificity of the HIV western blot and antibody tests for HIV and hepatitis C, any testing errors would be minimal. The observed difference in PPV and estimated specificity between the HIV and hepatitis C datasets could be attributed to (1) differences in the sensitivity and specificity of the underlying laboratory tests for HIV and HCV and (2) potentially greater rates of anonymous HIV testing, whereby public laboratories do not require full names for HIV testing [15].

Beyond the false-positive record linkages identified by examining the concordance of linked test results for HIV and hepatitis C, there is potential for other false-positives to occur in cases where individuals share common patient identifiers, such as twins. Given the deidentified nature of ACCESS data, without the actual identifying demographic values, these niche cases cannot be identified. The small impact of these false-positives is not expected to impact the main purpose of public health surveillance using ACCESS. For other research projects that require a lower level of false-positive record linkage, particularly if it is known to contain a high proportion of individuals sharing common patient identifiers, then using a linkage approach that only accepts linkage based on a match of multiple linkage keys would minimize false-positives. In addition, ensuring concordance of other extracted data, such as sex, year of birth, HIV, and hepatitis C testing history, can reduce the level of false-positive record linkages to acceptable levels.

Conclusions

Evaluating record linkage is an important part of assessing the utility of surveillance and research systems for answering key population-level research questions or for accurately describing population-level trends using linked data. Our findings suggest that the GRHANITE Linkage Tool is appropriate for accurately linking individuals' episodes of care and underpins the ability for ACCESS to perform privacy-preserving linkage of patient medical records.

Acknowledgments

The authors would like to acknowledge the contribution of the ACCESS Team members who are not coauthors of this study including: Lisa Bastian, WA Health; Deborah Bateson, Family Planning New South Wales (NSW); Scott Bowden, Doherty Institute; Mark Boyd, University of Adelaide; Allison Carter, Kirby Institute, University of New South Wales (UNSW) Sydney; Aaron Cogle, National Association of People with HIV Australia; Jane Costello, Positive Life NSW; Wayne Dimech, National Serology Reference Laboratory; Jennifer Dittmer, Burnet Institute; Jeanne Ellard, Australian Federation of AIDS Organisations; Christopher Fairley, Melbourne Sexual Health Centre; Lucinda Franklin, Victorian Department of Health; Jane Hocking, University of Melbourne; Jules Kim, Scarlet Alliance; Scott McGill, Australasian Society for HIV Medicine; David Nolan, Royal Perth Hospital; Prital Patel, Kirby Institute, UNSW Sydney; Stella Pendle, Australian Clinical Laboratories; Victoria Polkinghorne, Burnet Institute; Thi Nguyen, Burnet Institute; Catherine O'Connor, Kirby Institute, UNSW; Philip Reed, Kirkton Road Centre; Norman Roth, Prahran Market Clinic; Nathan Ryder, NSW Sexual Health Service Directors; Christine Selvey, NSW Ministry of Health; Michael Traeger, Burnet Institute; Toby Vickers, Kirby Institute, UNSW Sydney; Melanie Walker, Australian Injecting and Illicit Drug Users League; Lucy Watchirs-Smith, Kirby Institute, UNSW Sydney; Michael West, Victorian Department of Health.

The authors also acknowledge all clinics participating in ACCESS, PrEPX or both, and their site investigators for the provision of data to support these analyses: David Baker, East Sydney Doctors; Susan Boyd, Prahran Market Clinic; Mike Catton, Victorian Infectious Diseases Reference Laboratory; Danielle Collins, The Alfred; Vincent Cornelisse, Melbourne Sexual Health Centre; Pauline Cundill, PRONTO!; Philip Cunningham, SydPath, St Vincent's Pathology; Sian Edwards, Prahran Market Clinic; Christopher K Fairley and, Melbourne Sexual Health Centre; Robert Finlayson; Taylor Square; George Forgan-Smith, Era Health; John Gall, Era Health; Helan Lau, Prahran Market Clinic; Peter Locke, PRONTO! clinic; Anna McNulty, Sydney Sexual Health Centre; Richard Moore, Northside Clinic; Emma Paige, The Alfred; Matthew Penn, PRONTO! clinic; Claire Pickett, Ballarat Community Health Centre; William Rawlinson, South Eastern Area Laboratory Services, NSW Health Pathology; Norman Roth, Prahran Market Clinic; Hans-Gerhard Schneider, Alfred Health; BK Tee, The Centre Clinic; Amanda Wade, Geelong Hospital; Jeff Willcox, Northside Clinic.

The work contributed to this paper was in partnership with Treatment with Antiretrovirals and their Impact on Positive And Negative men (TAIPAN) and we would like to acknowledge members of the TAIPAN study who are not coauthors of this paper including: Andrew Carr, St Vincent's Hospital Sydney; Jennifer Hoy, Alfred Health; Kathy Petoumenos, Kirby Institute, UNSW; Julian Elliot, Alfred Health; David Templeton, Kirby Institute; Teng Liaw, School of Public Health and Community Medicine, UNSW; David Wilson, Burnet Institute; Andrew Grulich, Kirby Institute, UNSW; David Cooper, Kirby Institute, UNSW; Alisa

Pedrana, Monash University; James McMahon, Alfred Health; Garrett Prestage, Kirby Institute; Martin Holt, Centre for Social Research in Health, UNSW; Christopher K. Fairley, Melbourne Sexual Health Centre; Neil McKellar-Stewart, ACON Health Sydney; Simon Ruth, Thorne Harbour Health, Phillip Keen, Kirby Institute; Craig Cooper, Positive Life NSW; Brent Allan, Living Positive Victoria; John Kaldor, Kirby Institute, UNSW.

The authors would also like to acknowledge the PrEPX Study team including the Principal Investigator Edwina Wright, Alfred Health and study coordinators: Brian Price; Luxi Lal; John T. Lockwood; Anne Mak; Christina Chang; Judith Armishaw, Alfred Health.

ACCESS is a partnership between the Burnet Institute, Kirby Institute and National Reference Laboratory. ACCESS is funded by the Australia Department of Health. ACCESS also receives funding from specific studies, including EC Victoria, EC Australia, and PrEPX. The Burnet Institute gratefully acknowledges support from the Victorian Operational Infrastructure Support Program.

Conflicts of Interest

MH and MS have received investigator-initiated funding from Gilead Sciences, AbbVie, and Bristol Myers Squibb for research unrelated to this work. MH, RG, MS, and BD are supported by fellowships from the National Health and Medical Research Council.

References

1. Callander D, Moreira C, El-Hayek C, Asselin J, van Gemert C, Watchirs Smith L, et al. Monitoring the control of sexually transmissible infections and blood-borne viruses: protocol for the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance (ACCESS). JMIR Res Protoc 2018 Nov 20;7(11):e11028 [FREE Full text] [doi: 10.2196/11028] [Medline: 30459142]
2. van Gemert C, Guy R, Stoove M, Dimech W, El-Hayek C, Asselin J, et al. Pathology laboratory surveillance in the Australian Collaboration for Coordinated Enhanced Sentinel Surveillance of sexually transmitted infections and blood-borne viruses: protocol for a cohort study. JMIR Res Protoc 2019 Aug 8;8(8):e13625 PMCID: PMC6786847. [doi: 10.2196/13625]
3. Australian Government. Department of Health. 2018. Third National Hepatitis B Strategy 2018-2022 URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/ohp-bbvs-1//$File/Hep-B-Third-Nat-Strategy-2018-22.pdf [accessed 2019-10-21]
4. Australian Government. Department of Health. 2018. Fourth National Sexually Transmissible Infections Strategy 2018-2022 URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/ohp-bbvs-1//$File/STI-Fourth-Nat-Strategy-2018-22.pdf [accessed 2019-10-21]
5. Australian Government. Department of Health. 2018. Eighth National HIV Strategy 2018-2022 URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/ohp-bbvs-1//$File/HIV-Eight-Nat-Strategy-2018-22.pdf [accessed 2019-10-21]
6. Boyle D, Kong F. A systematic mechanism for the collection and interpretation of display format pathology test results from Australian primary care records. Electron J Health Inform 2011;6(2):e18 [FREE Full text]
7. Liaw S, Boyle D. Secure data linkage and information sharing with GRHANITE. In: Grain H, editor. HIC 2008 Conference: Australias Health Informatics Conference; The Person in the Centre, August 31 - September 2, 2008 Melbourne Convention Centre. Australia: Health Informatics Society of Australia Ltd (HISA); 2008:159-165.
8. Randall SM, Ferrante AM, Boyd JH, Bauer JK, Semmens JB. Privacy-preserving record linkage on large real world datasets. J Biomed Inform 2014 Aug;50:205-212 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.003] [Medline: 24333482]
9. Boyle DIR, Rafael N. BioGrid Australia and GRHANITE: privacy-protecting subject matching. Stud Health Technol Inform 2011;168:24-34. [Medline: 21893908]
10. Boyle DI, Versace VL, Dunbar JA, Scheil W, Janus E, Oats JJ, MAGDA Study Group. Results of the first recorded evaluation of a national gestational diabetes mellitus register: challenges in screening, registration, and follow-up for diabetes risk. PLoS One 2018;13(8):e0200832 [FREE Full text] [doi: 10.1371/journal.pone.0200832] [Medline: 30089149]
11. University of Melbourne HaBIC Research Information Technology Unit. Using GRHANITE Hashes as generated by GRHANITE Client March 2017 V4.0. Melbourne: University of Melbourne HaBIC Research Information Technology Unit; 2017. URL: http://www.grhanite.com/downloads/Using_GRHANITE_Hashes_generated_by_GRHANITE_Client_V4_0.pdf [accessed 2020-04-03]
12. Zhu Y, Matsuyama Y, Ohashi Y, Setoguchi S. When to conduct probabilistic linkage vs deterministic linkage? A simulation study. J Biomed Inform 2015 Aug;56:80-86 [FREE Full text] [doi: 10.1016/j.jbi.2015.05.012] [Medline: 26004791]
13. Ryan KE, Mak A, Stoove M, Price B, Fairley CK, Ruth S, et al. Protocol for an HIV Pre-exposure Prophylaxis (PrEP) Population Level Intervention Study in Victoria Australia: The PrEPX Study. Front Public Health 2018;6:151 [FREE Full text] [doi: 10.3389/fpubh.2018.00151] [Medline: 29896468]
14. Emergency Care Research Institute.: ECRI Institute; 2016. Patient Identification Errors. Health Technology Assessment Information Service, Special Report URL: https://www.ecri.org/Resources/HIT/Patient%20ID/Patient_Identification_Evidence_Based_Literature_final.pdf [accessed 2019-10-21]

15. ASHM Testing Portal. 2011 Nov. National HIV Testing Policy v1.2 URL: http://testingportal.ashm.org.au/resources/Australian_National_HIV_Testing_Policy_v1-2.pdf [accessed 2019-10-21]
16. Harron K, Goldstein H, Wade A, Muller-Pebody B, Parslow R, Gilbert R. Linkage, evaluation and analysis of national electronic healthcare data: application to providing enhanced blood-stream infection surveillance in paediatric intensive care. PLoS One 2013;8(12):e85278 [FREE Full text] [doi: 10.1371/journal.pone.0085278] [Medline: 24376874]

Abbreviations

ACCESS: Australian Collaboration for Coordinated Enhanced Sentinel Surveillance
BBV: blood-borne virus
EMR: electronic medical record
NSW: New South Wales
PPV: positive predictive value
STI: sexually transmissible infection
TAIPAN: Treatment with Antiretrovirals and their Impact on Positive And Negative men
UNSW: University of New South Wales

Edited by G Eysenbach; submitted 21.10.19; peer-reviewed by D Bandara, K Bosh, N Sundar Rajan, K Goniewicz; comments to author 03.12.19; revised version received 25.01.20; accepted 21.02.20; published 24.06.20.

Please cite as:
Nguyen L, Stoové M, Boyle D, Callander D, McManus H, Asselin J, Guy R, Donovan B, Hellard M, El-Hayek C
Privacy-Preserving Record Linkage of Deidentified Records Within a Public Health Surveillance System: Evaluation Study
J Med Internet Res 2020;22(6):e16757
URL: https://www.jmir.org/2020/6/e16757
doi:10.2196/16757
PMID:32579128

Original Paper

Association of eHealth Literacy With Lifestyle Behaviors in University Students: Questionnaire-Based Cross-Sectional Study

Saki Tsukahara[1*], MS; Satoshi Yamaguchi[1,2*], MD, PhD; Futaba Igarashi[1], MS; Reiko Uruma[3], MD, PhD; Naomi Ikuina[3], RN; Kaori Iwakura[3], PHN; Keisuke Koizumi[1], PhD; Yasunori Sato[4], PhD

[1] Collage of Liberal Arts and Sciences, Chiba University, Chiba, Japan

[2] Department of Orthopaedic Surgery, Graduate School of Medical and Pharmaceutical Sciences, Chiba University, Chiba, Japan

[3] Safety and Health Organization, Chiba University, Chiba, Japan

[4] Clinical and Translational Research Center, Keio University Hospital, Tokyo, Japan

[*] these authors contributed equally

Corresponding Author:
Satoshi Yamaguchi, MD, PhD
Collage of Liberal Arts and Sciences
Chiba University
1-33, Yayoi-cho, Inage-ku
Chiba, 263-8522
Japan
Phone: 81 43 251 1111
Email: y-satoshi@mvb.biglobe.ne.jp

Abstract

Background: Maintenance of good health and a healthy lifestyle have significant impacts on the lives of university students. However, university students are prone to engage in risky health behaviors, resulting in impaired health status. Electronic health (eHealth) literacy is an important factor in maintaining a healthy lifestyle. However, no studies have assessed the eHealth literacy levels and the associated lifestyle behaviors among university students in Japan.

Objective: The purposes of this study were to clarify the eHealth literacy level, the participant characteristics associated with eHealth literacy, and the association of eHealth literacy with lifestyle behaviors of students in a Japanese university.

Methods: A questionnaire-based cross-sectional study of 3183 students at a national university in Japan was conducted. eHealth literacy was quantified using the Japanese version of the eHealth Literacy Scale (eHEALS). The association between participant characteristics (gender, school year, department of study, and living status) and eHEALS score was assessed using *t* tests. Additionally, the associations of eHealth literacy with lifestyle behaviors (exercise, smoking, alcohol consumption, etc.) were evaluated using logistic regression analyses.

Results: The mean eHEALS score was 23.6/40 points. The mean eHEALS score for students in medical departments was 27.0/40 points, which was 2.9 points higher than that of nonmedical students ($P<.001$). Similarly, the graduate school participants had higher scores than the undergraduate students. The proportion of participants who exercised regularly was higher in the high eHEALS score group than in the low score group, with an adjusted odds ratio of 1.39 ($P<.001$).

Conclusions: The eHealth literacy level of university students in Japan was comparable to that of the general Japanese population. Graduate students, as well as those in medical departments, had higher eHealth literacy. Furthermore, students with higher eHealth literacy had better exercise routines.

(J Med Internet Res 2020;22(6):e18155) doi:10.2196/18155

KEYWORDS

college student; ehealth literacy; ehealth; eHealth Literacy Scale; health literacy; lifestyle; university student

Introduction

Healthy lifestyle behaviors, such as exercising regularly, sleeping well, and eating breakfast, have a significant impact on university student life. For example, students with healthy lifestyles achieve higher academic degrees than those without [1,2]. Moreover, an unhealthy lifestyle during university life, including smoking and alcohol use, tends to persist after graduation and is highly likely to create risk of lifestyle-related diseases in later life [3,4]. However, university students are prone to engage in risky health behaviors [5]. The transition from high school to university involves a drastic change in lifestyle, as students often start living away from their parents, developing new social networks, and having more free time than before [5]. Accordingly, a decline in university students' health status is associated with these lifestyle changes [6-8]. Therefore, the maintenance of a healthy lifestyle is a challenge for university students.

Health literacy is defined as an individual's knowledge, motivation, and skills to access, understand, evaluate, and apply health information [9]. Studies have shown that people with high health literacy have healthy lifestyles [10]. Therefore, improving health literacy has been identified as a public health goal and a significant health care challenge globally [11]. Currently, health information is often obtained through the internet, especially by members of the younger generation, including university students [12,13]. Collecting data through the internet is different from collecting data through books and leaflets, and it requires specific skills; people must not only be health literate but must also be able to find, understand, and appraise information using digital services and technology [14]. Accordingly, electronic health (eHealth) literacy—the ability to search for, evaluate and use health information on the internet to solve health problems—is considered to be an essential factor for university students to maintain a healthy lifestyle and good health status [15]. According to the Lily model [16], eHealth literacy consists of six core skills (or literacies), which are classified into two types: analytic skills (traditional, media, and information) and context-specific skills (computer, scientific, and health). Therefore, health literacy is one of the literacies that comprise the concept of eHealth literacy [16]. The eHealth Literacy Scale (eHEALS) is an 8-item questionnaire that was developed based on the Lily model and measures a broad overview of self-perceived eHealth literacy skills [12]. The eHEALS has been translated globally and is widely used to assess eHealth literacy levels [14]. Studies have evaluated the eHealth literacy levels of university students in different countries [15]. However, the eHealth literacy level can vary among countries. Specifically, the eHealth literacy level of Japanese people may be lower than those of people in European countries [17]. The eHealth literacy level of students in Japanese universities has not been studied.

In the general population, many personal and social background characteristics, including age, gender, household income, educational level, and occupation, are associated with health literacy levels [17,18]. Among university students, male gender, higher school year, medical study department, higher academic achievement, and higher family income are associated with higher eHealth literacy levels [19-26]. However, these studies have limitations, such as relatively small numbers of participants [19,21,23,26] and recruitment of student participants from only medical and nursing departments [21,24,25]. Therefore, a study with a larger number of participants, including both nonmedical and medical students, is required to clarify the relationship between student characteristics and eHealth literacy.

Studies have shown that people with higher eHealth literacy have healthier lifestyles than those with lower eHealth literacy in the general population [18]. Regarding university students, only a few studies have reported on the association between eHealth literacy and lifestyle behaviors, such as regular exercise, healthy eating, and regular sleep, in Taiwan, Greece, and the United States [20,22,27]. However, the numbers of participants in these studies [20,22,27] were relatively small. Furthermore, the influence of eHealth literacy on lifestyle behavior may vary depending on the country, cultural background, and degree of internet use [17,19,21].

The purposes of this study were to clarify the eHealth literacy level, the participant characteristics associated with eHealth literacy, and the association of eHealth literacy with lifestyle behaviors of students in a Japanese university.

Methods

Recruitment

This study was a questionnaire-based cross-sectional study performed at Chiba University, Japan. Chiba University is a national university with 13,983 students at the time of the study. Of those, 5306/13,983 (37.9%) were female and 8677/13,983 (62.1%) were male. Furthermore, 10,547/13,983 (75.4%) were undergraduate students, and 2430/13,983 (17.4%) students were studying medical sciences, including medicine, nursing, and pharmacy. Inclusion criteria were students who underwent on-campus medical examinations from April to May 2019. Exclusion criteria were students who declined to participate and who did not understand the Japanese questionnaire. Furthermore, students with incomplete answers to the questionnaire were excluded. We recruited participants during 12/19 checkup days. Of 13,983 university students, 5310 (38.0%) underwent the checkup during the 12 days. Of those, 1918/5310 (36.1%) declined to participate, and the remaining 3392 students (63.9%) answered the questionnaire. No student was excluded because of inability to understand the Japanese questionnaire. After excluding 209 students with incomplete answers, the data from 3183 students were used for analysis. The Chiba University Ethics Committee approved this study (approval number 01-02). The data were collected anonymously. No gifts or payments were given to participants for participating in this study.

eHealth Literacy Level

The questionnaire consisted of questions on eHealth literacy, participant characteristics, and lifestyle behaviors.

The Japanese version of the eHealth Literacy Scale (eHEALS) was used to quantify eHealth literacy [12,28]. Both the original and Japanese versions of the scale have sufficient reliability and validity to evaluate the eHealth literacy of adults [12,28]. The eHEALS consisted of eight questions, including four items related to internet capability and four items related to the utilization of health information [12]. Answer options were provided with a 5-point scale from "I totally do not think so (1)" to "I think so quite (5)" for each item. The total score of eight items was calculated, with a higher score indicating a better eHealth literacy level. The participants were divided into two groups using the mean eHEALS score for statistical analysis: a high score group (≥24, n=1659) and a low score group (<23, n=1524) .

Participant Characteristics

The participant characteristics included gender, school year, department of study, and living status. Answers for school year (first to sixth year of undergraduate studies and first to fourth year of graduate school) were dichotomized into undergraduate and graduate [20]. Answers for department of study (undergraduate: education, engineering, science, horticulture, humanities, public affairs, law, politics and economics, liberal arts and sciences, literature, medicine, nursing, pharmacy; graduate: education, horticulture, medical and pharmaceutical sciences, nursing, science and engineering, and law school) were divided into medical (undergraduate: medicine, nursing, pharmacy; graduate: medical and pharmaceutical science and nursing) and nonmedical [19,21]. Answers for living status (living alone, living with parents or partner, living in a dormitory, and other) were classified into living alone and living with others.

Lifestyle

Lifestyle behaviors were assessed using the questions on exercise, breakfast, smoking, alcohol consumption, and hours of sleep. Answers for exercise frequency (≥3 days/week, 1-2 days/week, 1-2 days/month, none) were dichotomized into ≥1 day/week and <1 day/week for statistical analysis [18]. Answers for breakfast (every day, 5-6 days/week, 1-4 days/week, none) were divided into ≥5 days/week and <5 days/week [29]. For smoking, the answers for the three choices (yes, no, previously) were divided into nonsmoker (no, previously) and current smoker (yes) [30]. Answers concerning alcohol consumption (do not drink, 1-2 days/month, 1-2 days/week, ≥3 days/week) were categorized into <3 days/week and ≥3 days/week [31]. Answers for hours of sleep (≤6 hours, 7-8 hours, ≥9 hours) were dichotomized into sufficient (7-8 hours, ≥9 hours) and insufficient (<7 hours) [29]. Additionally, the participants' BMI (kilograms per square meter) was recorded as a proxy of lifestyle. The answers were provided with either underweight (<18.5), normal (≥18.5, <25), overweight (≥25, <30), or obese (≥30) [32]. Overweight and obese were combined into one "overweight" category [32]. Although the answer for BMI was self-reported, we assume it was accurate because the participants' BMIs were measured during the health examination immediately before they participated in the questionnaire survey.

Statistical Analysis

Demographic data on the participants' characteristics and lifestyle were expressed using descriptive statistics. Numbers and frequencies were used for categorical variables. Means and standard deviations were used for continuous variables because most of the data had a normal distribution. Participants were dichotomized into two groups depending on the characteristic (eg, undergraduate/graduate). The eHEALS scores were compared between groups using Student *t* tests. Furthermore, the associations of eHEALS scores with lifestyle behaviors were assessed using logistic regression analysis. The explanatory variable was the eHEALS score group (low/high), and the objective variables were lifestyle behaviors. The odds ratio (OR) was calculated for a healthy lifestyle in the high eHEALS group; the low score group served as the reference. Unadjusted ORs and ORs adjusted for participant characteristics (gender, school year, department of study, and living status) were expressed. Bonferroni corrections were conducted to adjust for multiplicity. Accordingly, statistical significance was set at *P*<.013 for the association between participant characteristics and eHEALS score and at *P*<.007 for the association between eHEALS score and lifestyle behavior.

Results

Participant Characteristics

Of the 3183 participants, 878 (27.6%) were female, and 2549 (80.1%) were undergraduate students (Table 1). Students in medical departments comprised 346/3183 (10.9%) of the survey participants.

For lifestyle behaviors, 1757/3183 (55.2%) participants exercised ≥1 day/week (Table 2). Only 140/3183 (4.3%) were current smokers, and 166 (5.2%) drank alcohol ≥ 3 days/week. Although 281/3183 (8.8%) participants were overweight, 480 (15.1%) were classified as underweight.

Table 1. Participant characteristics (N=3183).

Characteristic	Participants, n (%)
Gender	
Male	2305 (72.4)
Female	878 (27.6)
School year	
Undergraduate (n=2549, 80.1%)	
1	606 (19.0)
2	622 (19.5)
3	608 (19.1)
4	613 (19.3)
5	45 (1.4)
6	55 (1.7)
Graduate (n=634, 19.9%)	
1	320 (10.1)
2	271 (8.5)
3	28 (0.9)
4	15 (0.5)
Department of study	
Medical (n=346, 10.9%)	
Medicine	606 (19.0)
Nursing	55 (1.7)
Pharmacy	125 (3.9)
Graduate school of medical and pharmaceutical science	14 (0.4)
Graduate school of nursing	13 (0.4)
Nonmedical[a]	2837 (89.1)
Living status	
Living alone	1600 (50.2)
Living with others (n=1583, 49.7%)	
Living with parents	1530 (48.1)
Living in dormitory	46 (1.4)
Other	7 (0.2)

[a] The 2837 nonmedical students were studying education (n=436, 15.4%), engineering (n=887), science (n=257, 31.3%), horticulture (n=68, 2.4%), law, politics, and economics (n=402, 14.2%), liberal arts and science (n=101, 3.6%), literature (n=179, 6.3%), graduate education (n=28, 1.0%), graduate horticulture (n=7, 0.2%), graduate humanities and studies on public affairs (n=34, 1.2%), graduate science and engineering (n=427, 15.1%), and law (n=11, 0.4%).

Table 2. Lifestyle behaviors of the participants (N=3183).

Lifestyle behavior	Participants, n (%)
Exercise	
≥1 day/week (n=1757, 55.2%)	
≥3 days/week	657 (20.6)
1-2 days/week	1100 (34.6)
<1 day/week (n=1426, 44.8%)	
1-2 days/month	632 (19.9)
None	794 (24.9)
Breakfast	
≥5 days/week (n=2132, 67.0%)	
Every day	1651 (51.9)
5-6 days/week	481 (15.1)
< 5 days/week (n=1051, 33.0%)	
1-4 days/week	557 (17.5)
None	494 (15.5)
Smoking	
Nonsmoker (n=3043, 95.6%)	
No	2967 (93.2)
Previously	76 (2.4)
Smoker	140 (4.4)
Alcohol	
<3 days/week (n=3017, 96.1%)	
None	1208 (38.0)
1-2 days /month	1189 (37.4)
1-2 days/week	620 (19.5)
≥3 days/week	166 (5.2)
Sleep	
Sufficient (n=1878, 59.0%)	
7-8 hours	1822 (57.2)
≥9 hours	56 (1.8)
Insufficient (≤6 hours)	1305 (41.0)
BMI (kilograms per square meter)	
Normal (≥18.5, <25)	2422 (76.1)
Underweight (<18.5)	480 (15.1)
Overweight (≥25, n=281, 8.8%)	
≥25, <30	230 (7.2)
≥30	51 (1.6)

eHealth Literacy Level

The mean eHEALS score was 23.6/40 points (SD 6.8). The mean scores for each item ranged from 2.7 to 3.1. The lowest score for was obtained for Q6: “I have the skills I need to evaluate the health resources I found on the internet,” and the highest score was obtained for Q8: “I feel confident in using information from the Internet to make health decisions.”

Association Between Participant Characteristics and eHEALS Score

The mean eHEALS score for medical students was 2.9 points higher than that for nonmedical students (*P*<.001, Table 3). The eHEALS score was higher for graduate students than for undergraduate students (*P*=.003). However, the difference between the groups was 0.9 points. There was no difference between female and male gender (*P*=.18) or between students living alone and living with others (*P*=.02).

Table 3. Association between participant characteristics and eHEALS score (N=3183).

Characteristic (n)	eHEALS[a] score, mean (SD)	*P* value
Gender		
Male (2305)	23.6 (7.0)	.18
Female (878)	23.3 (6.3)	
School year		
Undergraduate (2549)	23.4 (6.8)	.003
Graduate (634)	24.3 (6.6)	
Department of study		
Medical (346)	27.0 (6.6)	<.001
Nonmedical (2837)	23.1 (6.7)	
Living status		
Living alone (1600)	23.8 (6.9)	.02
Living with others (1583)	23.3 (6.7)	

[a] eHEALS: eHealth Literacy Scale.

Association of eHEALS Score With Lifestyle

Overall, participants in the high eHEALS score group had a healthier lifestyle than those in the low score group (Table 4). In the high score group, 984/1659 students (59.3%) exercised regularly (≥1 day/week), while in the low score group, 773/1524 (50.7%) exercised regularly. The adjusted OR for the high score group was 1.39 (*P*<.001, Table 4). Also, 1141/1659 (68.8%) and 991/1524 (65.0%) participants in the high and low score groups, respectively, had breakfast regularly (adjusted OR 1.24; *P*=.007). Interestingly, the risk of being overweight was higher in the high score group (adjusted OR 1.49; *P*<.001). However, the number of overweight students was low in both groups, namely 176/1659 (10.6%) in the high score group and 105/1524 (6.9%) in the low score group.

Table 4. Association of eHEALS score with lifestyle (N=3183). OR values are for the high eHEALS score group (n=1659) relative to the low score group (n=1524).

Lifestyle behavior	Unadjusted OR[a] (95% CI)	*P* value	Adjusted[b] OR (95% CI)	*P* value	Model *P* value
Regular exercise	1.42 (1.23-1.63)	<.001	1.39 (1.21-1.61)	<.001	<.001
Regular breakfast	1.18 (1.02-1.37)	.02	1.24 (1.06-1.45)	.007	<.001
No smoking	1.24 (0.88-1.74)	.22	1.18 (0.8-1.67)	.36	<.001
Alcohol <3 days/week	0.80 (0.59-1.10)	.17	0.82 (0.59-1.12)	.21	<.001
Sufficient sleep	1.10 (0.96-1.27)	.17	0.91 (0.79-1.06)	.22	<.001
Overweight (n=2703)[c]	1.58 (1.23-2.05)	<.001	1.49 (1.20-2.02)	<.001	<.001
Underweight (n=2902)[d]	0.93 (0.76-1.13)	.47	.94 (0.77-1.15)	.54	.14

[a] OR: odds ratio.

[b] Adjusted for gender, school year, department of study, and living status.

[c] n=1421 (52.6%) and n=1282 (47.4%) in the high and low eHEALS score groups, respectively.

[d] n=1476 (50.9%) and n=1426 (49.1%) in the high and low eHEALS score groups, respectively.

Discussion

Principal Findings

We showed that the average eHEALS score of students at a Japanese national university was approximately 24 points out of 40. Several personal background characteristics, including school year and department of study, were associated with a high eHEALS score. Additionally, students with higher eHEALS scores demonstrated better exercise behaviors. To the best of our knowledge, this is the first such study of Japanese students and one of the largest studies to clarify the eHealth literacy levels and related lifestyle behaviors of university students. Our results provide important information to help university students improve their eHealth literacy and achieve healthier lifestyles.

eHealth Literacy Level

In this study, the mean eHEALS score of the participants was 23.6/40 points. This value is comparable to that of general Japanese adults, whose mean score was 23.5 points [28,29]. However, studies from other countries have reported higher scores. For example, the mean eHEALS score of 192 students at an Iranian medical and health science university was 28.2 points [21]. In a study of 422 American college students, most of whom were undergraduate students, the mean score was 31.9 [27]. The eHEALS scores of the general population in several Asian, American, and European countries were also relatively high, ranging from 28.1-30.5 [33-36]. In addition to the lower eHealth literacy level of Japanese people, the general health literacy level, measured by using the European Health Literacy Survey Questionnaire (HLS-EU-Q47) [37], was lower in Japan than in European countries [17]. Therefore, the health literacy level in Japan appears to be low not only in university students but in Japanese people in general. There are some potential explanations for the lower health literacy of Japanese people. First, accessible and understandable public health information sites, such as the National Institutes of Health website in America [38] and the National Health Service website in Britain [39], are not available in Japan [17]. Therefore, Japanese people may have difficulty acquiring health literacy. Poor accessibility to health information can affect the eHealth literacy levels of college students in other countries as well [21]. Second, surveys by the Japanese Cabinet Office and the Ministry of Education, Culture, Sports, Science and Technology suggest that students and younger people have lower self-esteem in Japan than in other countries [40]. Because we assessed self-reported health literacy levels, the students who participated in this study may have rated their health information skills lower than the actual levels [41]. Further research is needed to clarify the cause of low health literacy in Japanese university students. Additionally, our results indicate that there is significant room for improvement in the eHealth literacy of Japanese university students [15].

Association Between Participant Characteristics and eHEALS Score

In this study, undergraduate and graduate students in the medical sciences (ie, medicine, nursing, and pharmacy) had higher eHEALS scores than those in nonmedical departments. This result was consistent with a study of 566 Taiwanese college students, in which medical students had higher eHealth literacy in all dimensions than nonmedical students [20]. A higher eHealth literacy level in medical students was also observed in a survey of 192 Iranian university students [21]. These results are understandable because medical students are more exposed to medical and health information in their curriculum.

The graduate student participants had higher eHEALS scores than undergraduate students, although the difference in the mean score was only 0.9 points. In studies of medical and nursing students, a higher school year was associated with higher eHealth literacy [19,21,24]. Additionally, a survey of 630 Danish university students showed higher health literacy among students in master's degree programs than among those in bachelor's degree programs [26]. Our study showed a similar association between school year and eHealth literacy in a broader range of student populations, including medical and nonmedical students. A possible explanation for our result is that eHealth literacy levels improve during university life. Another explanation is that students with higher eHealth literacy are selected for admission to graduate programs. Further longitudinal studies are necessary to clarify this issue. Combining the results of this study and those of previous studies, education programs to improve eHealth literacy should focus on early-degree and nonmedical students.

Association of eHEALS Score With Lifestyle

This study showed that participants in the high eHEALS score group exercised more frequently than those in the low score group. This association was significant after adjusting for participant characteristics. A relationship between higher eHealth literacy and better exercise behaviors is consistently found in Taiwanese, American, and Greek university students [20,22,27] as well as in adult internet users in Japan [29]. In this survey, 1426/3138 participants (44.8%) exercised <1 day/week. Promoting regular exercise to this population would be a fundamental part of healthy lifestyle promotion. The results of this study suggest that eHealth literacy education improves exercise behaviors.

In this study, a higher eHEALS score was associated with regularly eating breakfast. Our result was in line with previous studies, which showed that higher eHealth literacy was correlated with healthy diet behavior among American and Taiwanese college students [20,27]. People who did not eat breakfast reported suboptimal dietary behavior, such as unhealthy food choices and eating snacks [42]. Furthermore, skipping breakfast is associated with several lifestyle-related diseases [43,44] and lower academic achievement in college students [2]. Therefore, regularly eating breakfast serves as an indicator of an overall healthy lifestyle [42]. However, breakfast skipping is prevalent among young adults [45] and university students [46]. In this study, 1051/3183 (33.0%) of participants ate breakfast less than five times a week. Therefore, enhancing eHealth literacy could promote regular breakfast eating among university students.

Smoking and excessive alcohol were not associated with eHEALS score. Our results were consistent with general surveys of Japanese adults, in which the eHEALS score [29] and the general health literacy level [47,48] were not associated with smoking or alcohol consumption but were associated with exercise and balanced nutrition. Similarly, no association of eHealth literacy with smoking or alcohol was found in Greek university students [22]. This may be because the rates of habitual drinkers (166/3183, 5.2%) and current smokers (140/3183, 4.4%) were low in this study; thus, it is difficult to assess the association of eHealth literacy with alcohol and smoking [48]. Indeed, percentages of habitual drinkers and smokers have declined among young Japanese people in the past decade [49]. Another possible reason is that other environmental factors, such as friends, independent living, and family history, may affect the drinking and smoking behaviors of university students more than their eHealth literacy [50].

Limitations

This study has several limitations. First, this study was conducted at a single national university. Therefore, the results may not apply to university students from other backgrounds. For example, the type of university (ie, public or private) could affect the students' eHealth literacy levels [19]. Although approximately 3400 students from a wide range of school years and departments were enrolled in this study, further studies that include multiple universities are needed. Second, the participants of this study may not represent the general Chiba University student population. For example, only 878/3183 (27.6%) of participants were female, while 38% of Chiba University students are female. Third, we recruited participants who had undergone health checkups. People who undergo health checkups are more health-conscious and have healthier lifestyles than those who do not [51]; the eHealth literacy level may have been even lower if students who did not undergo health examinations had participated in the survey. Fourth, several participant characteristics that can be associated with eHealth literacy, such as academic achievement [22], family income [22], educational level of parents [26], student health history [26], and internet skills [19], were not surveyed because of the practicality limits of administering the questionnaire at the health examination. Fifth, because this study was designed as a cross-sectional study, the causal relationship of eHealth literacy with lifestyle was not clarified. Future studies that evaluate the effects of eHealth literacy education are needed to confirm the causality. Finally, we used eHEALS, which was developed in 2006, to quantify the eHealth literacy levels of the participants. However, the utilization of the internet has changed significantly since the scale was developed. Specifically, social media and mobile devices are among the most popular ways to use the internet among younger people [52]. Therefore, although this scale is a valid measurement and has been used widely, it may not fully represent the eHealth literacy levels of university students [53].

Conclusions

The eHealth literacy level of Chiba University students was comparable to that of the general Japanese population. Graduate students, as well as those in medical departments, had higher eHealth literacy levels. Furthermore, the students with higher eHealth literacy levels demonstrated better exercise behaviors. Interventions to address eHealth literacy could help improve students' lifestyles, although further research is warranted.

Conflicts of Interest

None declared.

References

1. Farokhzadian J, Karami A, Azizzadeh Forouzi M. Health-promoting behaviors in nursing students: is it related to self-efficacy for health practices and academic achievement? Int J Adolesc Med Health 2018 Jun 28:Preprint. [doi: 10.1515/ijamh-2017-0148] [Medline: 29953403]
2. Burrows TL, Whatnall MC, Patterson AJ, Hutchesson MJ. Associations between Dietary Intake and Academic Achievement in College Students: A Systematic Review. Healthcare (Basel) 2017 Sep 25;5(4):60 [FREE Full text] [doi: 10.3390/healthcare5040060] [Medline: 28946663]
3. Lolokote S, Hidru TH, Li X. Do socio-cultural factors influence college students' self-rated health status and health-promoting lifestyles? A cross-sectional multicenter study in Dalian, China. BMC Public Health 2017 May 19;17(1):478 [FREE Full text] [doi: 10.1186/s12889-017-4411-8] [Medline: 28525995]
4. Jao NC, Robinson LD, Kelly PJ, Ciecierski CC, Hitsman B. Unhealthy behavior clustering and mental health status in United States college students. J Am Coll Health 2019;67(8):790-800. [doi: 10.1080/07448481.2018.1515744] [Medline: 30485154]
5. Almutairi KM, Alonazi WB, Vinluan JM, Almigbal TH, Batais MA, Alodhayani AA, et al. Health promoting lifestyle of university students in Saudi Arabia: a cross-sectional assessment. BMC Public Health 2018 Sep 05;18(1):1093 [FREE Full text] [doi: 10.1186/s12889-018-5999-z] [Medline: 30185167]
6. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. J Affect Disord 2013 May 15;148(1):1-11. [doi: 10.1016/j.jad.2012.11.026] [Medline: 23246209]
7. Ridner SL, Newton KS, Staten RR, Crawford TN, Hall LA. Predictors of well-being among college students. J Am Coll Health 2016;64(2):116-124. [doi: 10.1080/07448481.2015.1085057] [Medline: 26630580]
8. Wang X, Jia X, Zhu M, Chen J. Linking health states to subjective well-being: an empirical study of 5854 rural residents in China. Public Health 2015 Jun;129(6):655-666. [doi: 10.1016/j.puhe.2015.03.014] [Medline: 25931439]
9. Organization for Economic Co-operation and Development. 2019 Dec 10. Health literacy for people-centered care URL: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DELSA/HEA/WD/HWP(2018)4&docLanguage=En [accessed 2020-06-03]
10. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]

11. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA 1999 Feb 10;281(6):552-557. [Medline: 10022112]
12. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
13. Escoffery C, Miner KR, Adame DD, Butler S, McCormick L, Mendell E. Internet use for health information among college students. J Am Coll Health 2005;53(4):183-188. [doi: 10.3200/JACH.53.4.183-188] [Medline: 15663067]
14. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowl Manag ELearn Int J 2015;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
15. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011 Dec 01;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
16. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
17. Nakayama K, Osaka W, Togari T, Ishikawa H, Yonekura Y, Sekido A, et al. Comprehensive health literacy in Japan is lower than in Europe: a validated Japanese-language assessment of health literacy. BMC Public Health 2015 May 23;15:505 [FREE Full text] [doi: 10.1186/s12889-015-1835-x] [Medline: 26001385]
18. Suka M, Odajima T, Okamoto M, Sumitani M, Igarashi A, Ishikawa H, et al. Relationship between health literacy, health information access, health behavior, and health status in Japanese people. Patient Educ Couns 2015 May;98(5):660-668. [doi: 10.1016/j.pec.2015.02.013] [Medline: 25739344]
19. Tubaishat A, Habiballah L. eHealth literacy among undergraduate nursing students. Nurse Educ Today 2016 Jul;42:47-52. [doi: 10.1016/j.nedt.2016.04.003] [Medline: 27237352]
20. Yang S, Luo Y, Chiang C. The Associations Among Individual Factors, eHealth Literacy, and Health-Promoting Lifestyles Among College Students. J Med Internet Res 2017 Jan 10;19(1):e15 [FREE Full text] [doi: 10.2196/jmir.5964] [Medline: 28073739]
21. Dashti S, Peyman N, Tajfard M, Esmaeeli H. E-Health literacy of medical and health sciences university students in Mashhad, Iran in 2016: a pilot study. Electron Physician 2017 Mar;9(3):3966-3973 [FREE Full text] [doi: 10.19082/3966] [Medline: 28461871]
22. Vozikis A, Drivas K, Milioris K. Health literacy among university students in Greece: determinants and association with self-perceived health, health behaviours and health risks. Arch Public Health 2014;72(1):15 [FREE Full text] [doi: 10.1186/2049-3258-72-15] [Medline: 24987522]
23. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
24. Park H, Lee E. Self-reported eHealth literacy among undergraduate nursing students in South Korea: a pilot study. Nurse Educ Today 2015 Feb;35(2):408-413. [doi: 10.1016/j.nedt.2014.10.022] [Medline: 25466791]
25. Newsome C, Mallawaarachchi I, Conklin J, Ray G. Health literacy of student pharmacists. Curr Pharm Teach Learn 2018 Mar;10(3):340-343. [doi: 10.1016/j.cptl.2017.11.005] [Medline: 29764638]
26. Elsborg L, Krossdal F, Kayser L. Health literacy among Danish university students enrolled in health-related study programmes. Scand J Public Health 2017 Dec;45(8):831-838. [doi: 10.1177/1403494817733356] [Medline: 29052485]
27. Britt RK, Collins WB, Wilson K, Linnemeier G, Englebert AM. eHealth Literacy and Health Behaviors Affecting Modern College Students: A Pilot Study of Issues Identified by the American College Health Association. J Med Internet Res 2017 Dec 19;19(12):e392 [FREE Full text] [doi: 10.2196/jmir.3100] [Medline: 29258979]
28. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. Developing Japanese version of the eHealth Literacy Scale (eHEALS). Article in Japanese. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
29. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
30. Aaby A, Friis K, Christensen B, Rowlands G, Maindal HT. Health literacy is associated with health behaviour and self-reported health: A large population-based study in individuals with cardiovascular disease. Eur J Prev Cardiol 2017 Nov;24(17):1880-1888 [FREE Full text] [doi: 10.1177/2047487317729538] [Medline: 28854822]
31. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
32. Expert Consultation W. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan 10;363(9403):157-163. [doi: 10.1016/S0140-6736(03)15268-3] [Medline: 14726171]
33. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): Reliability and Validity Testing in Younger Adults Recruited Online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]

34. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
35. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the Hungarian version of the eHealth Literacy Scale. Eur J Health Econ 2019 Jun;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]
36. Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, et al. Correlation Between eHealth Literacy and Health Literacy Using the eHealth Literacy Scale and Real-Life Experiences in the Health Sector as a Proxy Measure of Functional Health Literacy: Cross-Sectional Web-Based Survey. J Med Internet Res 2018 Oct 31;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
37. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, HLS-EU Consortium. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 2013 Oct 10;13:948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
38. National Institute of Health. URL: https://www.nih.gov/
39. National Health Service. URL: https://www.nhs.uk/ [accessed 2020-06-03]
40. Ministry of Education, Culture, Sports, Science and Technology, Japan. Low self-esteem among Japanese children. Report in Japanese URL: https://www.kantei.go.jp/jp/singi/kyouikusaisei/chousakai/dai1/siryou4.pdf [accessed 2020-06-03]
41. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006 Apr 21;8(2):e6 [FREE Full text] [doi: 10.2196/jmir.8.2.e6] [Medline: 16867969]
42. Uzhova I, Peñalvo JL. Reply: Skipping Breakfast Is a Marker of Unhealthy Lifestyle. J Am Coll Cardiol 2018 Feb 13;71(6):708-709 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.066] [Medline: 29420974]
43. Takagi H, Hari Y, Nakashima K, Kuno T, Ando T, ALICE (All-Literature Investigation of Cardiovascular Evidence) Group. Meta-Analysis of Relation of Skipping Breakfast With Heart Disease. Am J Cardiol 2019 Sep 15;124(6):978-986. [doi: 10.1016/j.amjcard.2019.06.016] [Medline: 31326073]
44. Bi H, Gan Y, Yang C, Chen Y, Tong X, Lu Z. Breakfast skipping and the risk of type 2 diabetes: a meta-analysis of observational studies. Public Health Nutr 2015 Nov;18(16):3013-3019. [doi: 10.1017/S1368980015000257] [Medline: 25686619]
45. Pendergast FJ, Livingstone KM, Worsley A, McNaughton SA. Correlates of meal skipping in young adults: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 01;13(1):125 [FREE Full text] [doi: 10.1186/s12966-016-0451-1] [Medline: 27905981]
46. Almutairi KM, Alonazi WB, Vinluan JM, Almigbal TH, Batais MA, Alodhayani AA, et al. Health promoting lifestyle of university students in Saudi Arabia: a cross-sectional assessment. BMC Public Health 2018 Sep 05;18(1):1093 [FREE Full text] [doi: 10.1186/s12889-018-5999-z] [Medline: 30185167]
47. Ishikawa H, Nomura K, Sato M, Yano E. Developing a measure of communicative and critical health literacy: a pilot study of Japanese office workers. Health Promot Int 2008 Sep;23(3):269-274. [doi: 10.1093/heapro/dan017] [Medline: 18515303]
48. Goto E, Ishikawa H, Nakayama K, Kiuchi T. Comprehensive Health Literacy and Health-Related Behaviors Within a General Japanese Population: Differences by Health Domains. Asia Pac J Public Health 2018 Nov;30(8):717-726. [doi: 10.1177/1010539518806806] [Medline: 30466315]
49. Endo J, Yamaguchi S, Saito M, Morikawa T, Akagi R, Sasho T. Changes in the syndesmotic reduction after syndesmotic screw fixation for ankle malleolar fractures: One-year longitudinal evaluations using computer tomography. Injury 2016 Oct;47(10):2360-2365. [doi: 10.1016/j.injury.2016.07.031] [Medline: 27469402]
50. Lorant V, Nicaise P. Binge drinking at University: a social network study in Belgium. Health Promot Int 2015 Sep;30(3):675-683. [doi: 10.1093/heapro/dau007] [Medline: 24622535]
51. Chien SY, Chuang MC, Chen IP. Why People Do Not Attend Health Screenings: Factors That Influence Willingness to Participate in Health Screenings for Chronic Diseases. Int J Environ Res Public Health 2020 May 17;17(10):E3495 [FREE Full text] [doi: 10.3390/ijerph17103495] [Medline: 32429532]
52. Information & Communications Statistics Database. Communication usage trend survey. Webpage in Japanese URL: http://www.soumu.go.jp/johotsusintokei/statistics/statistics05.html [accessed 2020-06-03]
53. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]

Abbreviations

eHEALS: eHealth Literacy Scale
eHealth: electronic health
HLS-EU-Q47: European Health Literacy Survey Questionnaire
OR: odds ratio

Edited by G Eysenbach; submitted 07.02.20; peer-reviewed by E Neter, L Kayser; comments to author 16.03.20; revised version received 01.04.20; accepted 16.04.20; published 24.06.20.

Please cite as:
Tsukahara S, Yamaguchi S, Igarashi F, Uruma R, Ikuina N, Iwakura K, Koizumi K, Sato Y
Association of eHealth Literacy With Lifestyle Behaviors in University Students: Questionnaire-Based Cross-Sectional Study
J Med Internet Res 2020;22(6):e18155
URL: http://www.jmir.org/2020/6/e18155/
doi:10.2196/18155
PMID:32579126

Original Paper

Disseminating the Foundations of Knowledge Translation and Patient Engagement Science Through the KnowledgeNudge Blog and Twitter Profile: Quantitative Descriptive Evaluation

Kathryn M Sibley[1,2], PhD; Masood Khan[1], MA; Patricia L Roche[2], MSc; Patrick Faucher[2], BA; Carly Leggett[2], MPH

[1]Department of Community Health Sciences, University of Manitoba, Winnipeg, MB, Canada

[2]George & Fay Yee Centre for Healthcare Innovation, Winnipeg, MB, Canada

Corresponding Author:
Kathryn M Sibley, PhD
Department of Community Health Sciences
University of Manitoba
379-753 McDermot Ave
Winnipeg, MB
Canada
Phone: 1 204 594 5354
Fax: 1 204 594 5394
Email: kathryn.sibley@umanitoba.ca

Abstract

Background: There is a documented need to build capacity for theory- and evidence-informed knowledge translation (KT) and patient engagement (PE) practice in health research. Dissemination of foundational content online coupled with social media promotion may build capacity by increasing awareness, knowledge, and positive attitudes.

Objective: This retrospective study sought to (1) describe exposure and engagement of the KnowledgeNudge KT and PE dissemination strategy (online blog and Twitter profile) over 2 years and (2) identify and compare characteristics of individual posts with the most and least exposure and reach.

Methods: Exposure was assessed by blog site views per month and Twitter profile impressions per month. Engagement was assessed by Twitter profile interactions per month. Descriptive statistics were calculated for 6-month blocks and compared using one-way analysis of variance or Student *t* test. Individual post exposure was assessed by average post views per week. Individual post reach was assessed by average post reads per week. High- and low-profile blog posts with the highest and lowest 10th percentile for exposure and reach were identified.

Results: A total of 99 posts and 755 tweets were published during the study period. There was a significant increase in exposure (P=.004) and reach (P<.001) during the final 6 months. Seven high-profile and 6 low-profile posts were identified. High-profile posts had a significantly greater average word count than low-profile posts (P=.003). There were no other significant differences between posts.

Conclusions: The increases in KnowledgeNudge exposure and engagement offer preliminary evidence in support of this dissemination strategy for the practice of KT and PE. Variation in individual post exposure and reach warrants further exploration to tailor content to user needs. Future work will include a prospective evaluation strategy to explore the effect of KnowledgeNudge on awareness, knowledge, attitudes, and behavior.

(J Med Internet Res 2020;22(6):e15351) doi:10.2196/15351

KEYWORDS

knowledge translation; dissemination; blog; Twitter

Introduction

Knowledge translation (KT) is an evolving discipline dedicated to advancing the synthesis, exchange, dissemination, and application of knowledge to optimize health, health care delivery, and the health care system [1]. It is a dynamic and iterative process that occurs throughout complex interactions between evidence producers and users [1], including but not limited to people with personal experience of a health issue (as a patient, family member, friend, or caregiver). The conceptual and evidence foundation for KT is growing rapidly, alongside growing recognition of the need for meaningful and active involvement of individuals with personal experience of a health issue in all stages of the research process (known in Canada as patient engagement [PE]) [2].

Accordingly, the need to build capacity to practice theory- and evidence-informed KT and PE has been identified in multiple studies [3-8]. For example, in a 2015 qualitative study of health researchers in Manitoba, Canada, KT education and training resources were identified as most needed to support practicing KT [7]. Similarly, in a 2017 survey of Manitoba health researchers, 81% of respondents indicated the need for training and educational resources to practice PE [9]. While the need for training has been consistently demonstrated, potential solutions need to consider known barriers to participation such as competing priorities and cost, which were identified in a 2014 Canadian study of knowledge translation training needs [10].

Dissemination of foundational resource material may help to build capacity for practicing theory- and evidence-informed KT and PE by facilitating distribution and awareness of seminal works, current best evidence and practices, and ongoing critical debates in the field [8]. In turn, dissemination may lead to increased knowledge and positive attitudes toward theory- and evidence-informed KT and PE, ultimately leading to improved skills and stimulation of behavior change [11]. Although effective dissemination strategies specific to KT and PE content have not been determined, online resources that use web-based technologies have emerged as a popular dissemination strategy due to their potential to reach large numbers of users [12] and have been shown to increase self-reported knowledge, skills, and information exchange in a time- and cost-efficient manner [13,14]. The impact of online dissemination resources may be enhanced when coupled with social media, which is rapidly growing in use in health research and care [15]. A 2015 study of 852 health researchers reported that 26.9% used social media for obtaining research evidence and almost all (95.9%) participants considered social media important for obtaining and disseminating research evidence [16]. Disseminating evidence through social media has also been shown to be effective in stimulating behavior change among health research users. For instance, a 2015 study of 317 clinicians found that 70% of participants reported changing or intending to change their practice after receiving evidence-based practice information through social media [17].

To address the established need to build capacity for practicing evidence-informed KT and PE, the KnowledgeNudge dissemination strategy was developed and launched in 2016. KnowledgeNudge consists of an online blog (a web-based collection of self-published content) [18] and accompanying Twitter profile [19]. The objective of KnowledgeNudge is to increase awareness, knowledge, and positive attitudes of KT and PE and related best practice concepts, considerations, and resources. The ultimate goal of KnowledgeNudge is to influence uptake of evidence-informed and/or best practice behaviors in KT and PE. Target audiences include health researchers, practitioners, and people with lived experience of a health issue. The editorial team comprises an academic KT scientist, KT and PE practice leads, and knowledge brokers with 4 to 10 years of experience. Contributors include a core group of KT scientists, academic trainees, research coordinators, KT and PE practice leads, and knowledge brokers (n=12), as well as guest posts from external health researchers and practitioners, patients/health consumers, and knowledge brokers with related expertise (n=9). Weekly blog posts summarize theoretical concepts, offer practice guidance, and provide opinions on issues of debate. Blog posts are intentionally short (less than 2000 words) and are written in a tone that is engaging, unique, and conversational and uses principles of plain language to appeal to a wide range of readers. Tweets are posted to promote each new blog post. Tweets direct readers to new content on the day of publication and through multiple follow-up posts in the first week. When new posts are not available, tweets promote existing posts.

There is consistent recognition of the need for outcome evaluation in established KT process models [20] and best practice recommendations for online KT tools and dissemination approaches [12]. As an important foundational step in evaluation of the KnowledgeNudge dissemination strategy, the overall goal of this study was to examine spread of KnowledgeNudge during its first 2 years. This concept was operationalized using recommended key performance indicators for social media use in health promotion: exposure, engagement, and reach [21]. Exposure reflects the number of times content has been viewed on a social media application. Engagement has been defined as an indicator of linking social medial to action, at its lowest level indicating some active interaction or participation, with individuals acknowledging agreement or preference for the content [21]. Reach indicates the number of individuals who have contact with the social media application and the related content [21].

The primary objective of this study was to describe and determine changes in exposure and engagement of the KnowledgeNudge blog and Twitter profile over a 2-year period (objective 1). The secondary objective was to identify and compare characteristics of individual posts with the most and least exposure and reach (objective 2). For the purpose of this study, exposure was measured through blog site views and Twitter profile impressions, and engagement was evaluated through Twitter profile interactions. Individual post exposure was evaluated through post views and reach through post reads.

Methods

Study Design

A retrospective descriptive quantitative study was conducted from August 1, 2016, through July 31, 2018. Given the nonidentifiable nature of the secondary data used for analysis, research ethics approval was not sought per Article 2.4 of the 2014 Government of Canada Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans [22].

Data Collection

For objective 1 (describe and determine changes in exposure and engagement of the KnowledgeNudge blog and Twitter profile), study data were extracted from standard use metrics available for the KnowledgeNudge blog site account (hosted on WordPress from August 2016 to February 2017 and on Medium from March 2017 to July 2018) and KnowledgeNudge Twitter profile. For objective 2 (identify and compare characteristics of individual posts with the most and least exposure and reach), metrics for individual posts were collected on October 1, 2018, allowing the last post in the study period (published on July 25, 2018) to accumulate metrics for 68 days. Individual post data were not available for the period August 2016 to February 2017 due to the change in site host and the inability to retrospectively retrieve individual post data from WordPress once the account was closed.

Outcomes

The primary study outcomes were exposure and engagement of the blog site and Twitter profile (objective 1), as well as individual post exposure and reach (objective 2) [23,24]. Exposure was defined as the total number of KnowledgeNudge blog site page views per month and the total number of Twitter profile impressions per month. Engagement was defined as the total number of Twitter profile engagements per month. Individual post exposure was defined as the average number of views per week, calculated for each post by dividing the total number of post views over the study period by the number of weeks since the post was first published. Individual post reach was defined as the average number of reads per week, calculated for each post by dividing the total number of reads for the study period (a metric provided by Medium, in which a read is counted when someone remains on the post page for the estimated amount of time it takes to read the entire post) by the number of weeks since the post was first published. Secondary outcomes included individual post content characteristics of post topic (knowledge translation, patient engagement, both, or other); post type (conceptual, practice, or other); when each post was published (dividing the study period into 6-month bins); the number of tweets used to share each post during the study period; and the total number of words per post (excluding the title, subtitle, and author information).

Data Analysis

Descriptive summary statistics were calculated for all variables (values reported are mean and standard deviation). To determine characteristics of individual posts, a combination of deductive and inductive coding was used. First, the research team met and developed a coding framework that consisted of two broad coding categories by topic (KT and PE) and three subcategories by type (practical issues or practice, theoretical concepts or conceptual, opinions, or other). Practice posts were defined as those providing information on tools and resources for KT and PE; conceptual posts were defined as those providing evidence-based information on the science of KT and PE based on peer-reviewed literature; and opinion posts were defined as those providing expert opinion on KT and PE science and practice. Two team members independently coded individual posts. An additional category was included for posts that specifically identified as relevant to both KT and PE, and all other posts that did not fit in the developed categories were coded as other (for example, informative posts about staff members). Coding disagreements were resolved through discussion. Kappa scores were calculated to determine interrater reliability of individual post coding, which was found to be strong for post topic (=.85) and moderate for post type (=.66) [25].

To determine changes in exposure and engagement over time (objective 1), a 1-way analysis of variance was conducted to assess average blog site exposure over time and Twitter profile exposure and engagement for each 6-month block. Post hoc comparisons were completed using the Tukey honestly significant difference (HSD) test. Statistical significance was set at $P<.05$.

To identify and compare individual posts with the most and least exposure and reach (objective 2), posts were ranked separately by exposure and reach (average weekly views and reads, respectively). High-profile posts were identified as those in the 10th highest percentile for both exposure and reach, and low-profile posts were identified as those in the 10th lowest percentile for both exposure and reach. Posts with missing data (those originally posted on WordPress and imported to Medium) were included in identification of high-profile posts but not in identifying low-profile posts in order to avoid categorizing these posts as low profile when exposure and reach could not be accurately measured for the entire study period. Tweets per post, word count, and exposure and reach of high- and low-profile posts were compared with a Student t test [26]. A 2-sided Fisher exact test was used to compare time of publication (by 6-month block), topic, and type of post for high- and low-profile posts [27].

Results

KnowledgeNudge Production Summary

A total of 99 posts were published during the study period (average 4 posts per month). Average post length was 818 words (range 246 to 1908). Of the total posts, 59% (58/99) were characterized as KT, 29% (29/99) as PE, 5% (5/99) as both KT and PE, and 7% (7/99) as other topics such as providing information about the KnowledgeNudge team. Half of the posts (49/99) were coded as practice, 31% (32/99) as conceptual, 12% (12/99) as opinion pieces, and 6% (6/99) as other. A total of 755 tweets were produced during the study period. Most tweets directly promoted individual posts (599/755, 79.3%), while the remainder promoted related content via retweets.

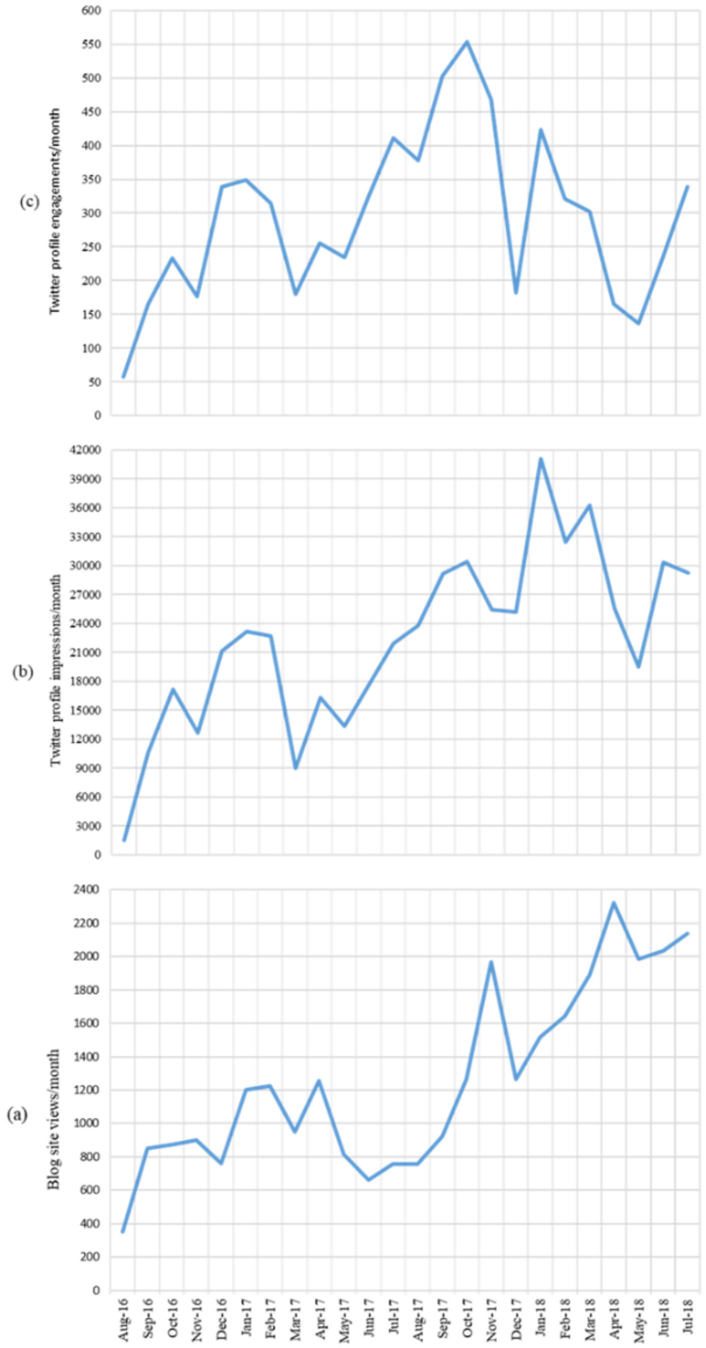
Objective 1. Exposure and Engagement of KnowledgeNudge

Monthly exposure and engagement are displayed in Figure 1. Average blog site exposure was 1263 (SD 549) site views per month during the study period (range 353 to 2322). There was a significant increase in blog site exposure over time ($F_{3,20}=17.9$, $P<.001$). Post hoc comparisons using the Tukey HSD test showed that average blog site exposure was significantly greater in the last 6-month block (2002 [SD 229] views per month) than the first (823 [SD 275] views per month) and second (944 [SD 248] views per month) 6-month blocks ($P=.001$) and the third (1283 [SD 432] views per month) 6-month block ($P=.003$). There was no significant difference in average blog site exposure between the first, second, and third 6-month blocks.

Average Twitter exposure was 22,320 (SD 9139) impressions per month (range 1522 to 41,112). A significant increase was seen in exposure over time ($F_{3,20}$=9.0, P<.001). Post hoc comparison using the Tukey HSD test showed that average Twitter exposure significantly increased from the first 6-month block (14,374 [SD 7903] impressions per month) to the third 6-month block (29,190 [SD 6371] impressions per month; P=.003) and to the last 6-month block (28,899 [SD 5780] impressions per month; P=.004). Post hoc comparison using the Tukey HSD test also revealed a significant increase in Twitter exposure between the second 6-month block (16,818 impressions per month), third 6-month block (P=.02), and last 6-month block (P=.02).

Average Twitter engagement was 294 (SD 124) engagements per month (range 57 to 554). There was a significant effect of time on average engagement ($F_{3,20}$=4.2, P=.02). Post hoc comparison using the Tukey HSD test showed a significant increase in average Twitter engagement between the first 6-month block (220 [SD 119] engagements per month) and the third 6-month block (418 [SD 131] engagements per month; P=.02). There were no other significant differences between the 6-month blocks.

Figure 1. KnowledgeNudge blog site exposure (a), Twitter exposure (b), and Twitter engagement (c) over 2-year period. There was a significant increase in blog site exposure (P<.001) and Twitter exposure (P=.004) between the first and last 6 months of the study period. No significant difference was seen in Twitter engagement between the first and last 6 months.



Objective 2: Individual KnowledgeNudge Post Characteristics

The average exposure of individual posts was 4.8 (SD 8.8) views per week (range 0.3 to 79). The average reach of individual posts was 1.7 (SD 1.7) reads per week (range 0.1 to 12). The average word count of individual posts was 988.3 (SD 429.3; range 434 to 1908). The average number of tweets per post was 7.1 (SD 2.3; range 3 to 11). Seven posts were identified as high profile, and 6 posts were identified as low profile. Comparison between high- and low-profile posts using a Student *t* test showed there was a significant difference between high- and low-profile posts for exposure (t_1=2.2, *P*=.02) and reach (t_1=4.9, *P*<.001). The word count of high-profile posts was significantly greater than that of low-profile posts (t_1=3.4, *P*=.003). There was no significant difference in the number of tweets per post between high- and low-profile posts (t_1=1.7, *P*=.06). There was no significant difference between high- and low-profile posts in terms of time of publication (*P*=.66), topic (*P*=.12), or type (*P*=.44). Characteristics of individual high- and low-profile posts are reported in Table 1.

Table 1. Characteristics of high- and low-profile KnowledgeNudge posts (those within the highest and lowest 10th percentile for both exposure and reach).

Title	Publication date	Topic	Type	Word count	Tweets/ post	Exposure (views/week)	Reach (reads/week)	Rank
The Knowledge-to-Action Framework	Nov 15, 2017	KT	Conceptual	1908	8	79	12	High
Decolonizing Community Engagement	Nov 22, 2017	PE	Conceptual	1411	11	38	9	High
Infographics: A Primer for Researchers	Jul 25, 2018	KT	Conceptual	1505	5	19	7	High
Knowledge Translation & Translational Research: Are They One & the Same?	Sep 6, 2017	KT	Conceptual	988	6	11	5	High
Budgeting for Patient & Public Engagement in Health Research	Oct 18, 2017	PE	Practice	1142	10	10	4	High
Unpacking KT Theories, Models, and Frameworks	Jan 25, 2017	KT	Conceptual	794	8	9	5	High
Top 10 Knowledge Translation Resources	Jul 12, 2017	KT	Practice	1072	8	9	4	High
Top KnowledgeNudge Posts of 2017	Nov 29, 2017	Other	Other	510	3	0.5	0.3	Low
Photovoice Blog-Series, Blog #2: Community Partnerships & Hard-to-Reach Populations	Aug 16, 2017	PE	Practice	915	4	0.9	0.4	Low
Everything (So Far) in One Post!	Apr 5, 2017	Other	Other	434	7	1	0.4	Low
Networks for Knowledge Translation	May 17, 2017	KT	Conceptual	554	9	1.4	0.8	Low
Sex- and Gender-Based Analysis (SGBA): Importance in Health Research	May 31, 2017	Both	Conceptual	640	7	1.6	0.6	Low
Part-II: The James Lind Alliance: An Overview of the Process of Priority Setting Partnerships	Apr 11, 2018	PE	Conceptual	975	6	1.8	0.7	Low

Discussion

Principal Findings

The goal of this study was to explore changes in exposure and engagement of the KnowledgeNudge KT and PE dissemination strategy during its initial 2-year period. By formally measuring and documenting changes in exposure and engagement of the KnowledgeNudge strategy, this study adds to the growing body of literature on the use of social media in dissemination science and demonstrates important contributions for building capacity in the practice of high-quality KT and PE practices. The findings establish a foundation for future studies to explore in-depth the impact of KnowledgeNudge content on KT and PE practice behaviors.

An important finding was the significant increase in blog site exposure between the first and last 6-month periods, providing evidence of increased spread of information on foundational issues of KT and PE. The significant increase in Twitter profile exposure during the study also provides evidence of growing network size. The absence of significant changes in Twitter profile engagement between the first and last 6-month periods is arguably less important, as the Twitter profile served to encourage users to visit the primary blog site and was not the primary source of KT and PE content. We also observed considerable variation in reach among the individual posts. Post topic and type coding did not suggest any trends influencing variation between posts with the highest and lowest reach.

The findings of this study are consistent with published evaluations of exposure and engagement in online dissemination blog strategies and Twitter activity in related domains of health [28-30]. The increase in exposure is a precursor to increase in awareness. The findings also support the goals of KnowledgeNudge by increasing the spread of information on foundational issues relating to the practice of KT and PE. A 2019 scoping review of core competencies for KT identified 19 individual competencies specific to addressing primarily knowledge, skills, or attitudes [31]. Although individual posts were not coded according to this framework, the KnowledgeNudge strategy as a whole seeks to address aspects of each of these competency domains. For example, the KnowledgeNudge dissemination strategy addresses many of these individual competencies related to the domain of knowledge including sharing of knowledge, increasing awareness of evidence resources, and promoting the understanding of research process, KT, and dissemination activities [31]. However, it has yet to be determined whether the KnowledgeNudge dissemination strategy addressed the domains of skills and attitudes. Moving forward, this core competency framework could be used to guide development of KnowledgeNudge content and evaluation approaches.

Although we did not directly assess factors that contributed to the increase in blog site exposure, using Twitter as a promotional strategy has been associated with increase in website traffic [29,32]. For example, a study of a website and Twitter profile aimed at disseminating resources related to child health found that an increase in Twitter exposure was associated with increased website reach [29]. Other factors that might have contributed to the increase in exposure of the blog site include posting content on a regular basis (on average one post per week) and publishing on subject matter identified as being of interest to readers through annual reader surveys. Similarly, having a team of authors with a wide of range of expertise also contributed to producing content covering a vast range of topics in PE and KT, potentially attracting a wide range of readership. Future studies should more explicitly consider and compare the effects of variables such as post frequency and content on changes in exposure, reach, and engagement.

The findings related to characteristics of high- and low-profile posts have important implications for those planning online dissemination strategies in general and those specifically interested in building capacity for KT and PE practices. Among the blog characteristics tested, the only significantly different variable between high- and low-profile posts was word count. Notably, high-profile posts had significantly more words than low-profile posts. This finding is supported by some evidence suggesting longer blogs are more influential in terms of garnering reader attention and feedback [33]. We found no difference in KT or PE post topic, practice or conceptual type, number of tweets per post, or time period of publication between high and low-profile posts. Other metrics not captured in our study may help to explain differences between high- and low-profile posts. The finding that time period of publication was not significantly different does eliminate two important opposite influencing factors: first, that older posts might accumulate more views and reads compared with newer posts because of the former being posted for a longer period on the website, and second, that more recent posts might have higher views and reads compared with older posts because of recency bias [34].

Limitations

We acknowledge the limitations of this retrospective study and potential for continued evaluation to overcome them. First, we recognize that reliance on online use metrics provides only a proxy indicator of awareness and generally relies on the assumption that the reader is actively interacting with the content. Analysis of related indicators such as Medium comments or claps (Medium’s version of likes) may have provided additional evidence for active engagement and awareness with the KnowledgeNudge strategy, although the inherent difficulty of accessing, measuring, and interpreting social media metrics is recognized in the literature [35]. Retrospective use of standard online metrics also restricted data accessibility to some extent (particularly blog site-level metrics such as unique visitors), and this was compounded by the unanticipated effects of changing host sites during the study period. We recommend those planning similar evaluations ensure in advance that their anticipated metrics are accessible, relevant, and feasible to collect.

Future research opportunities for understanding the effects of the KnowledgeNudge dissemination strategy include (but are not limited to) additional analysis to explore factors influencing variation in individual post engagement (such as author and guest effects; social media influencer effects; tweet content such as hashtags, photos, or language; and timing), social network analysis based on KnowledgeNudge Twitter and/or Medium followers, and a prospective study examining the overall impact of KnowledgeNudge on knowledge, skills, and attitudes toward theory- and evidence-informed KT and PE practice. Recognizing the limitations of dissemination strategies on behavior [36] and given that multifaceted KT strategies have been shown to be more effective [37] the impact of KnowledgeNudge would likely be optimized when coupled with active training and practice-building strategies.

Conclusions

The KnowledgeNudge dissemination strategy demonstrated increased online blog site and Twitter exposure over an initial 2-year period. Growth of KnowledgeNudge resulted mostly from an increase in blog site and Twitter exposures. The outcomes included in this study, based on standard use data, provide evidence of a foundational component for increasing capacity for theory- and evidence-informed KT and PE practice: access to information. Critical next steps include ongoing evaluation to explore the effects of KnowledgeNudge on knowledge, skills, and behavior. Variation in individual post reach warrants further exploration and audience feedback to tailor content to user needs. Future work will include a prospective evaluation strategy to comprehensively explore the effect of KnowledgeNudge on knowledge, attitudes, and behavior, and opportunities to leverage KnowledgeNudge resources in an active training program will be explored.

Acknowledgments

This study was funded by the George & Fay Yee Centre for Healthcare Innovation. KMS is supported through the Canada Research Chairs Program. We acknowledge the leadership of Kristy Wittmeier and Sane Dubé in establishing the KnowledgeNudge dissemination strategy and contributions of the many authors who have developed content for the blog.

Conflicts of Interest

None declared.

References

1. Canadian Institutes of Health Research. 2015. Guide to knowledge translation planning at CIHR: integrated and end-of-grant approaches URL: https://cihr-irsc.gc.ca/e/documents/kt_lm_ktplan-en.pdf [accessed 2020-05-12]
2. Canadian Institutes of Health Research. 2019. Strategy for patient-oriented research: patient engagement framework URL: http://cihr-irsc.gc.ca/e/48413.html#a4 [accessed 2020-05-12]
3. Eccles MP, Armstrong D, Baker R, Cleary K, Davies H, Davies S, et al. An implementation research agenda. Implement Sci 2009;4:18 [FREE Full text] [doi: 10.1186/1748-5908-4-18] [Medline: 19351400]
4. Gagliardi AR, Dobrow MJ. Identifying the conditions needed for integrated knowledge translation (IKT) in health care organizations: qualitative interviews with researchers and research users. BMC Health Serv Res 2016 Dec 12;16:256 [FREE Full text] [doi: 10.1186/s12913-016-1533-0] [Medline: 27405465]
5. Harvey G, Marshall RJ, Jordan Z, Kitson AL. Exploring the hidden barriers in knowledge translation: a case study within an academic community. Qual Health Res 2015 Nov;25(11):1506-1517. [doi: 10.1177/1049732315580300] [Medline: 25847856]
6. Michael Smith Foundation for Health Research Vancouver, BC. 2012. Training and resources to support research use: a BC needs assessment URL: https://www.msfhr.org/sites/default/files/KT_Needs_Assessment_FINAL.pdf [accessed 2020-05-12]
7. Sibley KM, Roche PL, Bell CP, Temple B, Wittmeier KDM. A descriptive qualitative examination of knowledge translation practice among health researchers in Manitoba, Canada. BMC Health Serv Res 2017 Sep 06;17(1):627 [FREE Full text] [doi: 10.1186/s12913-017-2573-9] [Medline: 28874152]
8. Birdsell JM, Omelchuk K. Alberta Heritage Foundation for Medical Research. 2007. Building capacity for health research transfer in western Canada: an environmental scan URL: http://www.assembly.ab.ca/lao/library/egovdocs/2007/alhfm/163517.pdf [accessed 2020-05-12]
9. Crockett LK, Shimmin C, Wittmeier KDM, Sibley KM. Engaging patients and the public in health research: experiences, perceptions and training needs among Manitoba health researchers. Res Involv Engagem 2019;5:28 [FREE Full text] [doi: 10.1186/s40900-019-0162-2] [Medline: 31608160]
10. Holmes BJ, Schellenberg M, Schell K, Scarrow G. How funding agencies can support research use in healthcare: an online province-wide survey to determine knowledge translation training needs. Implement Sci 2014 Jun 06;9:71 [FREE Full text] [doi: 10.1186/1748-5908-9-71] [Medline: 24906229]
11. Lafrenière D, Menuz V, Hurlimann T, Godard B. Knowledge dissemination interventions: a literature review. SAGE Open 2013 Aug 23;3(3):1-14. [doi: 10.1177/2158244013498242]
12. Levac D, Glegg SMN, Camden C, Rivard LM, Missiuna C. Best practice recommendations for the development, implementation, and evaluation of online knowledge translation resources in rehabilitation. Phys Ther 2015 Apr;95(4):648-662 [FREE Full text] [doi: 10.2522/ptj.20130500] [Medline: 25301966]
13. Curran VR, Fleet L. A review of evaluation outcomes of web-based continuing medical education. Med Educ 2005 Jun;39(6):561-567. [doi: 10.1111/j.1365-2929.2005.02173.x] [Medline: 15910431]
14. Mairs K, McNeil H, McLeod J, Prorok JC, Stolee P. Online strategies to facilitate health-related knowledge transfer: a systematic search and review. Health Info Libr J 2013 Dec;30(4):261-277. [doi: 10.1111/hir.12048] [Medline: 24251889]
15. Harseim T, Goodey G. Of Schemes and Memes. 2017 Jun 15. How do researchers use social media and scholarly collaboration networks (SCNs)? URL: http://blogs.nature.com/ofschemesandmemes/2017/06/15/how-do-researchers-use-social-media-and-scholarly-collaboration-networks-scns [accessed 2020-05-12]
16. Tunnecliff J, Ilic D, Morgan P, Keating J, Gaida JE, Clearihan L, et al. The acceptability among health researchers and clinicians of social media to translate research evidence to clinical practice: mixed-methods survey and interview study. J Med Internet Res 2015 May 20;17(5):e119 [FREE Full text] [doi: 10.2196/jmir.4347] [Medline: 25995192]
17. Maloney S, Tunnecliff J, Morgan P, Gaida JE, Clearihan L, Sadasivan S, et al. Translating evidence into practice via social media: a mixed-methods study. J Med Internet Res 2015 Oct 26;17(10):e242 [FREE Full text] [doi: 10.2196/jmir.4763] [Medline: 26503129]
18. KnowledgeNudge Blog. URL: https://medium.com/knowledgenudge [accessed 2018-08-03]
19. KnowledgeNudge Twitter Profile. URL: https://twitter.com/knowledgenudge [accessed 2018-08-03]

20. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
21. Neiger BL, Thackeray R, Van Wagenen SA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]
22. Canadian Institute of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: Ethical conduct for research involving humans. 2014. URL: http://www.frqnt.gouv.qc.ca/documents/10191/186009/TCPS2.+pdf/6a8ab915-431b-428d-aa86-b22ca5c78053 [accessed 2020-05-12]
23. About your activity dashboard: what is the Tweet activity dashboard?. URL: https://help.twitter.com/en/managing-your-account/using-the-tweet-activity-dashboard [accessed 2019-06-23]
24. Medium: your stats. URL: https://help.medium.com/hc/en-us/articles/215108608-Your-stats [accessed 2019-06-23]
25. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
26. Kalpic D, Hlupic N, Lovric M. Student's t-tests. In: Lovric M, editor. International Encyclopedia of Statistical Science. Springer: Berlin; 2011.
27. McDonald JH. Handbook of Biological Statistics, 3rd edition. Sparky House Publishing: Baltimore; 2014.
28. Carley S, Beardsell I, May N, Crowe L, Baombe J, Grayson A, et al. Social-media-enabled learning in emergency medicine: a case study of the growth, engagement and impact of a free open access medical education blog. Postgrad Med J 2018 Feb;94(1108):92-96. [doi: 10.1136/postgradmedj-2017-135104] [Medline: 29054933]
29. Gates A, Featherstone R, Shave K, Scott SD, Hartling L. Dissemination of evidence in paediatric emergency medicine: a quantitative descriptive evaluation of a 16-week social media promotion. BMJ Open 2018 Jun 06;8(6):e022298 [FREE Full text] [doi: 10.1136/bmjopen-2018-022298] [Medline: 29880576]
30. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014;16(7):e169 [FREE Full text] [doi: 10.2196/jmir.3430] [Medline: 25014311]
31. Mallidou AA, Atherton P, Chan L, Frisch N, Glegg S, Scarrow G. Core knowledge translation competencies: a scoping review. BMC Health Serv Res 2018 Dec 27;18(1):502 [FREE Full text] [doi: 10.1186/s12913-018-3314-4] [Medline: 29945609]
32. Dyson MP, Newton AS, Shave K, Featherstone RM, Thomson D, Wingert A, et al. Social media for the dissemination of Cochrane Child Health Evidence: evaluation study. J Med Internet Res 2017 Sep 01;19(9):e308 [FREE Full text] [doi: 10.2196/jmir.7819] [Medline: 28864427]
33. Agarwal N, Liu H, Tang L, Yu P. Identifying the influential bloggers in a community. Proc Int Conf Web Search Web Data Mining 2008 Feb:297-218 [FREE Full text] [doi: 10.1145/1341531.1341559]
34. Murphy J, Hofacker C, Mizerski R. Primacy and recency effects on clicking behavior. J Comp Mediated Comm 2006 Jan;11(2):522-535. [doi: 10.1111/j.1083-6101.2006.00025.x]
35. Drula G. Social and online media research: data, metrics and methods. 2012. URL: http://reaser.eu/RePec/rse/wpaper/10_Rev3_Drula_DigitalMedia.pdf [accessed 2020-05-12]
36. Gagnon ML. Moving knowledge to action through dissemination and exchange. J Clin Epidemiol 2011 Jan;64(1):25-31. [doi: 10.1016/j.jclinepi.2009.08.013] [Medline: 20004552]
37. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. Med Care 2001 Aug;39(8 Suppl 2):II2-I45. [Medline: 11583120]

Abbreviations

HSD: honestly significant difference
KT: knowledge translation
PE: patient engagement

Edited by G Eysenbach; submitted 03.07.19; peer-reviewed by A Cyr, S Ganesh, D Hansen; comments to author 21.10.19; revised version received 13.12.19; accepted 31.03.20; published 15.06.20.

Please cite as:
Sibley KM, Khan M, Roche PL, Faucher P, Leggett C
Disseminating the Foundations of Knowledge Translation and Patient Engagement Science Through the KnowledgeNudge Blog and Twitter Profile: Quantitative Descriptive Evaluation
J Med Internet Res 2020;22(6):e15351
URL: http://www.jmir.org/2020/6/e15351/
doi:10.2196/15351
PMID:32442133

Original Paper

Prediction of (Non)Participation of Older People in Digital Health Research: Exergame Intervention Study

Arianna Poli[1], MPsy; Susanne Kelfve[1,2], PhD; Leonie Klompstra[3], PhD; Anna Strömberg[3,4], Prof Dr; Tiny Jaarsma[3,5], Prof Dr; Andreas Motel-Klingebiel[1], Prof Dr

[1]Division Ageing and Social Change, Department of Culture and Society, Linköping University, Norrköping, Sweden
[2]Aging Research Center, Karolinska Institutet and Stockholm University, Stockholm, Sweden
[3]Department of Health, Medicine and Caring Science, Linköping University, Linköping, Sweden
[4]Department of Cardiology, Linköping University, Linköping, Sweden
[5]Julius Centre, University Medical Center, Utrecht, Netherlands

Corresponding Author:
Arianna Poli, MPsy
Division Ageing and Social Change
Department of Culture and Society
Linköping University
Campus Norrköping, Bredgatan 33
Norrköping, 60174
Sweden
Phone: 46 700850634
Email: arianna.poli@liu.se

Abstract

Background: The use of digital technologies is increasing in health care. However, studies evaluating digital health technologies can be characterized by selective nonparticipation of older people, although older people represent one of the main user groups of health care.

Objective: We examined whether and how participation in an exergame intervention study was associated with age, gender, and heart failure (HF) symptom severity.

Methods: A subset of data from the HF-Wii study was used. The data came from patients with HF in institutional settings in Germany, Italy, the Netherlands, and Sweden. Selective nonparticipation was examined as resulting from two processes: (non)recruitment and self-selection. Baseline information on age, gender, and New York Heart Association Functional Classification of 1632 patients with HF were the predictor variables. These patients were screened for HF-Wii study participation. Reasons for nonparticipation were evaluated.

Results: Of the 1632 screened patients, 71% did not participate. The nonrecruitment rate was 21%, and based on the eligible sample, the refusal rate was 61%. Higher age was associated with lower probability of participation; it increased both the probabilities of not being recruited and declining to participate. More severe symptoms increased the likelihood of nonrecruitment. Gender had no effect. The most common reasons for nonrecruitment and self-selection were related to physical limitations and lack of time, respectively.

Conclusions: Results indicate that selective nonparticipation takes place in digital health research and that it is associated with age and symptom severity. Gender effects cannot be proven. Such systematic selection can lead to biased research results that inappropriately inform research, policy, and practice.

Trial Registration: ClinicalTrial.gov NCT01785121, https://clinicaltrials.gov/ct2/show/NCT01785121

(J Med Internet Res 2020;22(6):e17884) doi:10.2196/17884

KEYWORDS

technology; exclusion; recruitment; self-selection; nonparticipation

Introduction

Increasingly more digital health technologies are evaluated for their potential use in the provision of health care services, with promising results at both individual and organizational levels [1-3]. It is expected that current health care service provision will be progressively restructured around the use of digital health technologies [4,5]. The use of digital technologies in health care service provision involves all users and hence, older people, as they represent one of the main user groups of health care services [6].

Although the number of older people using digital technologies is increasing [7], many are either nonusers or have little experience with using new technologies [8,9], even in countries that show high levels of internet and digital technology use [10].

Differences in the use of digital technologies exist not only between younger and older people but also within the older population [9]. Lower use and nonuse of digital technologies in old age are generally related to material, social, and individual resources [11]. Previous studies have found that lower education levels, less income, and poorer health status are associated with less use or nonuse of digital technologies [12-16]. Limited use or nonuse of digital technologies can also be determined by individual choices [17].

Such digital disparities can lead to disadvantages for some groups and hence contribute to increased risks of social exclusion and widened social inequality. Disparities can be reinforced by the participant selection strategies in studies evaluating new digital technologies [1,18,19]. Evaluations of digital health technologies typically follow the laboratory phases in which the digital technologies are developed and customized and focus on evaluating such tools with future target users via pilot trials or randomized trials [2]. However, such evaluations have some important shortcomings. Among others, the selection of study participants often results in samples that do not reflect the target population of interest [18,19] and hence, predetermines selective nonparticipation in the studies.

Selective nonparticipation in research prevents acknowledging, representing, and recognizing the interests and needs of different people [20,21], leading to biased research results [18,22] and generation of recommendations that are inaccurate for the entire population of interest. Underrepresented groups might be excluded or benefit less from the opportunities provided by the use of digital technologies in health care service provision and, most importantly, be disadvantaged by this change, compared with their digitally engaged counterparts. Such a disadvantage holds the potential to widen existing old-age inequalities [23,24].

Previous research has found that people who participate in studies on digital health technologies are typically younger, with better subjective health and overall quality of life; are more often men; are in better socioeconomic conditions; report more frequent digital access; and have higher digital skills [25-29]. Merkel et al [18] emphasized that vulnerable populations are less likely to be involved in such studies. In addition, refusal rates are often quite high [25,29,30]. Overall, people who agree to participate in studies can differ significantly from those who decline [29].

There are few studies that describe processes that result in selective nonparticipation of older people in digital health research [26,27,29]. Previous studies have mainly focused on individual decision making regarding whether to participate. However, less is known about selective nonparticipation resulting from the combination of two processes: (non)recruitment according to study requirements and self-selection as an individual decision.

The aim of this paper was to understand what predicts (non)participation of older people in digital health research. Here, participation is defined as being involved in digital health research as a study participant, and nonparticipation is defined as not being involved in digital health research. To study this, we conducted an analysis that deconstructs and examines (non)recruitment and self-selection occurring in a study evaluating exergame technology for patients with heart failure (HF); detailed descriptions of both the study and the intervention have been reported previously [31,32]. We investigated whether patients who were not eligible, those who declined, and those who accepted participation differed according to age, gender, and HF symptom severity. We hypothesized that age, gender, and symptom severity predicted individual participation in the HF-Wii study. We expected that participants would be younger, more likely to be men, and have less severe symptoms compared with those who were not eligible and those who declined to participate.

Methods

HF-Wii Study

This study was conducted within the research program “Supporting Self-Care by Information and Communication Technology for Older People with Long-Term Conditions” (ICT4Self-care; 2015-2018) conducted at Linköping University and funded by the Swedish Research Council for Health, Working Life and Welfare (FORTE; dnr 2014-4100). The analyses are based on data collected from the HF-Wii study [31,33], which contributed to this research program.

The HF-Wii study evaluated the impact of exergaming on the exercise capacity and daily physical activity of patients with HF. Exergame is a term that refers to video games that can be used for exercising, often at home. The HF-Wii study is based on a randomized controlled trial (RCT) approach. The HF-Wii study was developed based on the results obtained from a case study and a pilot study in which patients with HF were involved. Based on the experiences gained from these studies (ie, evaluating the results, experiences of patients with an exergame platform, and experiences of the research staff in conducting an exergame study), the RCT was developed [33,34]. Furthermore, a research partner (a patient with HF) from the Swedish Heart and Lung Foundation was involved in refining the instruction session, questionnaires, and recruitment and data collection strategies.

The target population of the HF-Wii study consisted of patients who were diagnosed with HF by a cardiologist according to

European Society of Cardiology guidelines [35] and who were older than 18 years. Exclusion criteria in the HF-Wii study were physical or balance problems, visual impairments, severe cognitive impairment(s) or psychiatric illness, a life expectancy shorter than 6 months, and not being able to speak or understand the language of the country where the study took place. Such criteria were assessed by a recruiter in each of the countries. Patients who were eligible according to the criteria were invited to participate in the HF-Wii study.

This study is based on data that refer to the study inclusion from four of the countries involved in the HF-Wii study: Germany, Italy, the Netherlands, and Sweden. Ethical approval for the HF-Wii study was obtained from local ethical committees (in Germany, GERS22(a)/2015; in Italy, IT:0052838/272/UVF/1; in the Netherlands, NL48647.068.14/METC141085; in Sweden, DNR 2012/247-31).

Data

The study population for this research in the four countries (Germany, Italy, the Netherlands, and Sweden) consisted of 1632 patients with HF who were screened in institutional settings between 2013 and 2017 for participation in the HF-Wii study. For these patients with HF, baseline information on age, gender, and HF symptom severity was completed by the recruiters in the four countries. Information on age was available for 1567 patients. Data on gender were retrieved for 1552 patients, and New York Heart Association (NYHA) class was recorded for 1180 patients.

Furthermore, data on (non)participation in the HF-Wii study based on the recruitment logs and reasons for nonparticipation were collected. No personal nor sensitive data were collected for this analysis.

Selective Nonparticipation

Selective nonparticipation was studied as resulting from two succeeding processes in the recruitment phase of the HF-Wii study: (non)recruitment according to the study requirements and self-selection as an individual decision [36]. The (non)recruitment process refers to the initial screening to select who is eligible to participate and excludes the others (ie, nonrecruited group). In contrast, self-selection is based on individual decision making that results in accepting or declining the invitation to participate and distinguishes between participants and decliners.

Outcome Variable

The outcome variable was constructed as a categorical variable with 3 categories, namely the 3 groups in which patients could alternatively be classified as a result of the recruitment phase: nonrecruited, decliner, or participant. The categories were constructed based on information from the recruitment logs that reported whether a patient was not recruited (ie, ineligible because of exclusion criteria), a decliner (ie, the patient declined the invitation to participate in the study), or a participant in the HF-Wii study.

Predictor Variables

Age, gender, and HF symptom severity were the predictor variables of this study. Age was included as a categorical variable: ≤64 years, 65-69 years, 70-79 years, ≥80 years, or missing. Gender was categorized as male, female, or missing. HF symptom severity was assessed by HF nurses according to the NYHA Functional Classification [37]. This classification is based on a subjective assessment of symptoms ranging between class I (ie, no symptoms and no limitation in ordinary physical activity, but presence of shortness of breath when walking or climbing stairs) and class IV (ie, severe limitations, experience of symptoms even while at rest). For the analyses, a missing value category was used.

Documented reasons for nonrecruitment and self-selection may further explain (non)recruitment and self-selection. Therefore, for those patients who were not eligible based on the HF-Wii study criteria, reasons for nonrecruitment were documented. For those patients who did not want to participate, reasons for declining the invitation were collected as free-text responses and coded into 10 categories: not having time, working or travelling a lot, unwilling to come to follow-up meetings, having other illnesses, already exercising a lot, unwilling to use technical equipment or the exergame device, already have a exergame device, living between different houses (unwillingness to move the exergame equipment from place to place), and shared living. Reasons for declining that did not fall into any of the above listed 10 categories were classified as “other.”

Analyses

First, descriptive analyses were performed to describe the predictor variables in the 3 groups and to illustrate nonrecruitment, declining, and participation rates. One-way ANOVA and Pearson’s chi-squared test were used for testing differences among the 3 groups.

Second, multinomial logistic regression was used to test the association between recruitment group membership (ie, nonrecruited, decliner, participant) and the predictors age, gender, and symptom severity. The participant group was the reference category.

Based on the regression model, we calculated the average marginal effects (AMEs) for each of the categories of the predictor variables. For each category of the predictor variables in the model, the AME showed the probability of being part of the nonrecruited, decliner, or participant groups for an individual who has the same values on every independent variable in the model except one [38]. AMEs were used because they could be compared more easily than odds ratios across groups in the sample [39].

In these analyses, we did not include patients for whom information on age and gender was missing. Inclusion of patients with missing data in the analyses was checked but did not result in a significant improvement of the model. However, we kept the missing value category for the variable NYHA as it concerns a larger number of cases. The sample size for these analyses consisted of 1489 patients (ie, net sample).

Third, to illustrate the selection processes, the reasons why patients were deemed ineligible, and why they decided not to participate in the HF-Wii study were evaluated. Reasons for ineligibility were investigated for all patients in the nonrecruited group, and reasons for individuals not to participate were described for patients in the decliner group.

Analyses were performed using SPSS Statistics version 25.0 (IBM Corp, Armonk, NY) and Stata software version 15 (StataCorp, College Station, TX).

Results

Patient Characteristics

Overall, 1632 patients with HF entered the recruitment phase of the HF-Wii study in the four countries. The mean age of the patients was 70 years (SD 11.9 years). More than half of the patients who entered the recruitment phase were ≥70 years old, representing 55.58% (907/1632) of the sample, and 13.30% (217/1632) of the patients were between 65 and 69 years of age. Around one quarter of the sample (443/1632, 27.14%) was ≤64 years old. No information on age was available for the remaining 3.98% (65/1632) of the patients. Of the 1632 patients screened, 64.83% (1058/1632) were men, and 30.27% (494/1632) were women. Information was missing for 4.90% (80/1632) of patients.

Of all the patients, 35.66% (582/1632) were classified as having mild HF symptoms and somewhat limited ability to exercise (NYHA class II), and 27.76% (453/1632) had marked limitations in activity due to symptoms (NYHA class III). Only 2.39% (39/1632) of patients had severe HF symptoms even at rest (NYHA class IV), and 6.50% (106/1632) had no HF symptoms (NYHA class I). Information on HF symptom severity was missing for 452 patients (452/1632, 27.70%).

Nonrecruitment, Declining, and Participation Rates

Overall, 71.45% (1166/1632) of all patients screened for the HF-Wii study did not participate in the intervention study (Table 1). Among those who did not participate, 37.99% (443/1166) were nonrecruited, and 62.01% (723/11166) declined to participate. The refusal rate for the HF-Wii study, based only on those patients who were invited to the study, was 60.81% (723/1189).

Mean ages were significantly different between the nonrecruited (73 years, SD 12.2 years), decliner (70 years, SD 11.8 years), and participant (67 years, SD 11.2 years) groups ($F_{2,1566}$=29.2, *P*<.001). Participants were significantly younger than patients in the decliner and nonrecruited groups.

Among the different age groups, the participation rate was lowest for patients 80 years old and older (59/361, 16.3%). The highest participation rate was found among patients 65-69 years of age (89/217, 41.0%). On the other hand, it was more common for patients 80 years old and older to be nonrecruited (155/361, 42.9%), compared with all the other groups (70-79 years, 127/546, 23.3%; 65-69 years, 40/217, 18.4%; and ≤64 years, 97/443, 21.9%). Declining to participate in the RCT was more common among patients 70-79 years of age (257/546, 47.1%), compared with the other age groups, for which declining varied from 40.6% to 42.9%.

On average, it was more common for men to participate in the HF-Wii study (332/1058, 30.38%) compared with women (134/494, 27.1%). In contrast, women were more often nonrecruited (168/494, 34.0%) than men (259/1058, 24.48%), while men more often declined participation (467/1058, 44.14%) compared with women (192/494, 38.9%). No information on gender was available for 16 patients in the nonrecruited group and for 64 patients in the decliner group.

On average, participation was lower among patients with marked limitations in activity due to HF symptoms and severe HF symptoms at rest (NYHA class III, 105/453, 23.2%; NYHA class IV, 4/39, 10.3%) compared with patients with no HF symptoms (NYHA class I, 55/106, 51.9%) or mild HF symptoms (NYHA class II, 280/582, 48.1%). Being ineligible for the HF-Wii study was more common among patients with NYHA class IV HF (28/39, 71.8%) than among the other groups for which the rate of nonrecruitment varied from 4.7% to 44.6%.

Table 1. Characteristics of the 1632 patients, by recruitment group.

Characteristics	Recruitment groups			Statistic	*P* value
	Nonrecruited, n (%)	Decliner, n (%)	Participant, n (%)		
Group size	443 (27.14)	723 (44.30)	466 (28.55)		
Age range (years)					
≤64	97 (21.90)	190 (42.89)	156 (35.21)	χ^2_8=113.4	<.001
65-69	40 (18.43)	88 (40.55)	89 (41.01)		
70-79	127 (23.26)	257 (47.07)	162 (29.67)		
≥80	155 (42.94)	147 (40.72)	59 (16.34)		
Missing	24 (36.92)	41 (63.08)	0 (0)		
Age (years)	73 (12.2)[a]	70 (11.8)[a]	67 (11.2)[a]	$F_{2,1556}$=29.2	<.001
Gender					
Men	259 (24.48)	467 (44.14)	332 (31.38)	χ^2_2=65.3	<.001
Women	168 (34.01)	192 (38.87)	134 (27.13)		
Missing	16 (20.00)	64 (80.00)	0 (0)		
HF[b] symptom severity					
NYHA class I[c]	5 (4.72)	46 (43.40)	55 (51.89)	χ^2_8=416.6	<.001
NYHA class II[d]	60 (10.31)	242 (41.58)	280 (48.11)		
NYHA class III[e]	202 (44.59)	146 (32.23)	105 (23.18)		
NYHA class IV[f]	28 (71.79)	7 (17.95)	4 (10.26)		
Missing	148 (32.74)	282 (62.39)	22 (4.87)		

[a] mean (SD).

[b] HF: heart failure.

[c] NYHA class I: no limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea (shortness of breath) [37].

[d] NYHA class II: slight limitation of physical activity. Comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea (shortness of breath) [37].

[e] NYHA class III: marked limitation of physical activity. Comfortable at rest. Less than ordinary activity causes fatigue, palpitation, or dyspnea [37].

[f] NYHA class IV: unable to carry on any physical activity without discomfort. Symptoms of HF at rest. If any physical activity is undertaken, discomfort increases [37].

Participation in the HF-Wii Study

A multinomial logistic regression was performed to test the association between recruitment group membership (ie, nonrecruited, decliner, participant) and the predictors age, gender, and HF symptom severity. The reference group consisted of patients who were recruited and agreed to participate (ie, participant group; Table 2).

The model shows that HF symptom severity, according to the NYHA functional classification, significantly predicted the likelihood of participation in the HF-Wii study. Patients with more severe symptoms, namely those in NYHA classes III and IV, had a decreased probability of participating and an increased probability of not being recruited, compared with patients with no or mild symptoms (ie, those in NYHA classes I and II). When compared with patients who displayed no symptoms (ie, NYHA class I), patients who had more severe HF symptoms (ie, NYHA classes III and IV) were significantly more likely to be nonrecruited to the HF-Wii study than to participate in it (NYHA class III, odds ratio [OR] 14.68, *P*<.001; NYHA class IV, OR 56.18, *P*<.001). Average marginal effects showed that more severe HF symptoms increased the probability of not being recruited by 63 percentage points for patients with NYHA class IV and by 35 percentage points for patients with NYHA class III (Figure 1). More severe HF symptoms decreased the probability of participation by 3 percentage points for patients with NYHA class I, by 25 percentage points for patients with NYHA class II, and by 39 percentage points for patients with NYHA class III.

Table 2. Multinomial logistic regression model of the relationship between recruitment group membership and predictors (ie, age, gender, and heart failure [HF] symptom severity) for all the patients for whom information on age and gender was available (n=1489) in the HF-Wii study. Pseudo R^2 (McFadden)=0.13.

	Nonrecruited vs participant[a]				Decliner vs participant[a]				Non recruited	Decliner	Participant
Predictors	β	SE	*P* value	Exp(β)	β	SE	*P* value	Exp(β)	AME[b] (SD)	AME (SD)	AME (SD)
Intercept	–2.93	.51	.00	N/A[c]	–0.64	.26	.01	N/A	N/A	N/A	N/A
Age range (years)											
≤64	.57	.26	.03	1.78	.29	.20	.15	1.34	.06 (.04)	.01 (.03)	–.08 (.04)
65-69 (ref.)	0	N/A	N/A	N/A	0	N/A	N/A	N/A	0	0	0
70-79	.46	.25	.07	1.59	.55	.20	.01	1.74	.02 (.03)	.08 (.04)	–.10 (.04)
≥80	1.35	.27	<.001	3.86	.89	.24	<.001	2.44	.13 (.04)	.06 (.04)	–.19 (.04)
Gender											
Women	.19	.16	.24	1.21	–.04	.14	.77	.96	.04 (.02)	–.03 (.03)	–.01 (.02)
Men (ref)	0	N/A	N/A	N/A	0	N/A	N/A	N/A	0	0	0
HF symptom severity[d]											
NYHA class I[e] (ref)	0	N/A	N/A	N/A	0	N/A	N/A	N/A	0	0	0
NYHA class II[f]	.70	.49	.15	2.02	.03	.23	.91	1.03	.05 (.03)	–.02 (.05)	–.03 (.05)
NYHA class III[g]	2.69	.49	<.001	14.68	.41	.25	.11	1.51	.35 (.04)	–.10 (.06)	–.25 (.06)
NYHA class IV[h]	4.03	.72	<.001	56.18	.67	.67	.32	1.95	.63 (.08)	–.24 (.08)	–.39 (.08)

[a] Reference category.

[b] AME: average marginal effect.

[c] N/A: not applicable.

[d] The missing value category was included in the analyses but is not reported in this table.

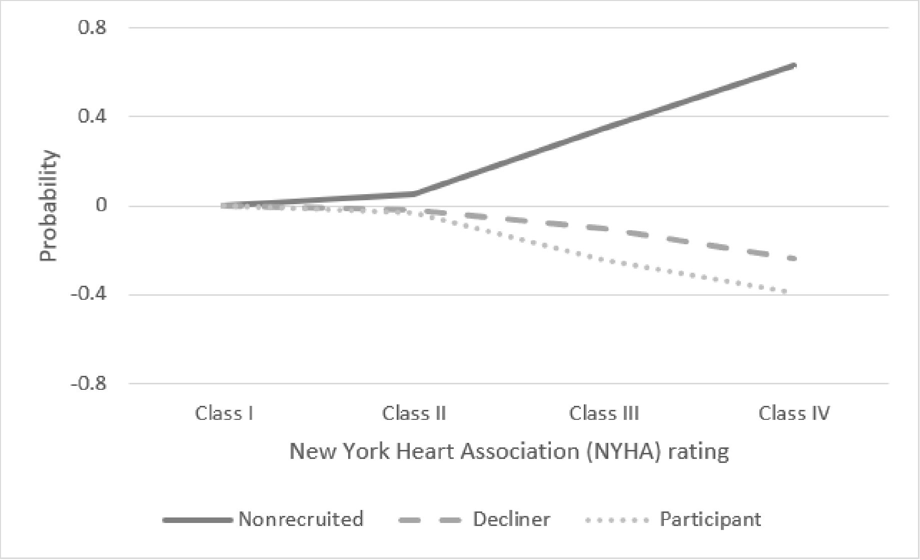
[e] NYHA class I: no limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea (shortness of breath) [37].

[f] NYHA class II: slight limitation of physical activity. Comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea (shortness of breath) [37].

[g] NYHA class III: marked limitation of physical activity. Comfortable at rest. Less than ordinary activity causes fatigue, palpitation, or dyspnea [37].

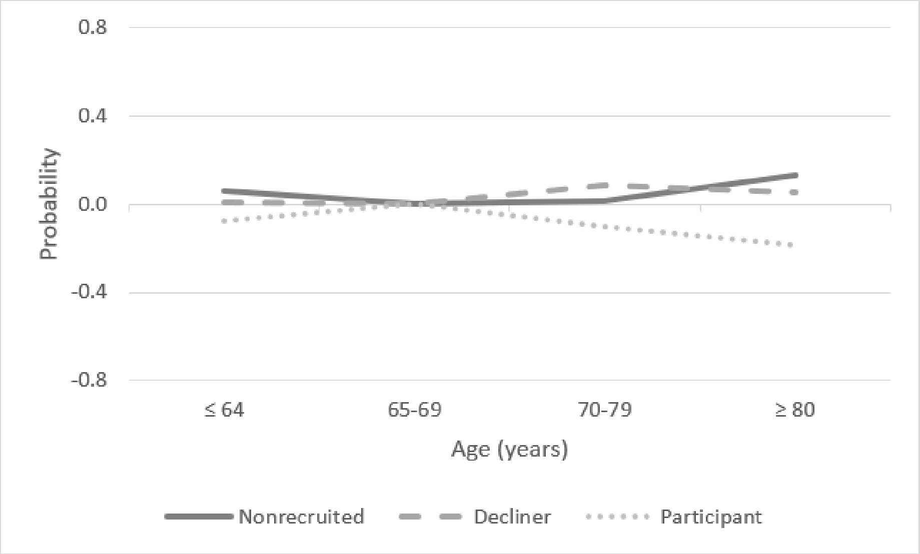
[h] NYHA class IV: unable to carry on any physical activity without discomfort. Symptoms of HF at rest. If any physical activity is undertaken, discomfort increases [37].

Figure 1. Probability of being in one of the 3 recruitment groups (ie, nonrecruited, decliner, participant) based on heart failure symptom severity.



The model also reveals that age was a significant predictor for participation in the HF-Wii study and contributed to explaining both (non)recruitment based on study requirements and self-selection as an individual decision. Being older reduced the probability of participating and of agreeing to participate. Compared with those 65-69 years of age, patients 80 years old and older were more likely to not be recruited than to be participants in the HF-Wii study (OR 3.86, *P*<.001). Probability of being not recruited increased by 13 percentage points for patients 80 years old and older, when compared with those 65-69 years of age (Figure 2). This was also found among patients 64 years of age and younger (OR 1.78, *P*=.03).

Figure 2. Probability of being in one of the 3 recruitment groups (ie, nonrecruited, decliner, participant) based on age.



Compared with the patients 65-69 years of age, people in the two older age groups (ie, 70-79 years and ≥80 years) were significantly more likely to decline the invitation to participate in the HF-Wii study than to participate in it (OR 1.74, *P*=.01; OR 2.44, *P*<.001, respectively). Being older increased the probability of declining participation in the study by 8 percentage points for patients 70-79 years of age and by 6 percentage points for those who were 80 years old and older, compared with patients 65-69 years of age.

Gender was not associated with the likelihood of participation in the HF-Wii study. It had no effect on the probability of being in the nonrecruited, decliner, or participant groups.

Reasons for Non-Recruitment and Declining to Participate

Of the patients, 27.14% (443/1632) were nonrecruited and thus were not invited to participate in the HF-Wii study. Patients were nonrecruited if they met at least one of the exclusion criteria.

The two most common reasons for being nonrecruited in the HF-Wii study were balance (223/443) and other physical problems (189/443; Table 3). Both reasons were more common among patients with marked and severe HF symptoms (ie, NYHA classes III and IV) than among the other patients. The least common reason for nonrecruitment was having a life expectancy <6 months (18/443). Age was a major factor for non-recruitment: 84% (187/223) of the nonrecruited patients with balance problems and 79% (149/189) of non-recruited patients with physical problems were 70 years old and older.

Table 3. Reasons for non-recruitment, based on the exclusion criteria and reasons for declining to participate in the HF-Wii study.

Reasons	n
Nonrecruitment (n=443)[a,b]	
Balance problems	223
Physical problems	189
Inability to fill in the questionnaire	107
Cognitive impairment(s)	79
Visual impairment	52
Life expectancy <6 months	18
Declining to participate (n=723)[c,d,e]	
Not having time	93
Working or travelling a lot	82
Unwilling to come to follow up meetings	48
Having other illnesses	41
Already exercising a lot	41

[a] Patients could be ineligible because of more than one criterion.

[b] Only the reasons given by at least 1 patient are reported.

[c] Multiple reasons for each patient were possible.

[d] Reasons were given by 569 patients (78.7%).

[e] Only reasons given by >30 patients are reported in the table.

Among the patients meeting the study criteria and invited to the HF-Wii study, 60.80% (723/1189) decided not to participate, and 79% percent (569/723) gave at least one reason for declining to participate (Table 3).

Among the reasons for declining, the 3 most often reported reasons were: “not having time,” “working or travelling a lot,” and “unwilling to come to follow-up meetings.” Not having time was the most common reason for declining participation in the HF-Wii study among patients 65-69 years of age and younger. It was more common among the youngest age group than among the other age groups to decline to participate in the HF-Wii study because they were already engaged in physical activities. Working or travelling a lot and not having time were the two most common reasons for declining to participate among patients 70 years and older.

Discussion

Evaluations of digital health technologies are affected by selective nonparticipation that can prevent the representation of needs and interests of parts of the target population, bias research results, and generate conclusions that lead to inefficient solutions and new inequalities. The aim of this paper was to understand the predictors of participation in digital health research. For this purpose, an analysis of selective nonparticipation in a study evaluating an exergame technology for patients with HF (ie, the HF-Wii study) was conducted to examine whether and how (non)participation is associated with age, gender, and symptom severity. Selective nonparticipation was deconstructed and analyzed as resulting from two consecutive processes: (non)recruitment based on study requirements and self-selection as the individual decision.

Overall, results show that participants, compared with nonparticipants, had less severe HF symptoms and were younger, but did not differ by gender. The main reasons for nonrecruitment were balance or physical problems, whereas the main reasons for declining participation in the HF-Wii study were related to lack of time and other commitments (ie, working or travelling a lot).

More severe HF symptoms predicted the nonrecruitment of patients in our sample. Patients with such severity of symptoms often show balance problems or severe comorbidities that more often make them ineligible according to study criteria. This result agrees with previous findings on participation in health-related research showing that patients with poorer health status, such as frail patients [40], patients with cognitive impairment [41], and patients with poorer physical functioning [42], are less likely to be participants.

However, among the patients in our sample, HF symptom severity did not significantly affect self-selection. This contrasts with findings of previous research describing individuals who decline participation as being more likely to show worse health than participants [43-45]. One possible explanation is that the patients with more severe HF symptoms were identified as ineligible to participate in the HF-Wii study already in the initial screening phase because of a higher incidence of balance or physical problems and, thus, were not invited to participate in the study.

Age significantly predicted participation in the evaluation of the exergame intervention. Belonging to an older age group reduced the probability of participating through both the processes of (non)recruitment and self-selection. This confirms previous findings on participation in health-related research

[46] and more specifically in both digital health [25,29] and HF research [47]. Such results are especially relevant in relation to the epidemiology of HF. As HF is more common among older people [48,49], it is crucial that the inclusion of older people in digital health research targeting HF is sustained.

Counter to expectations, patients 64 years old and younger were less likely to participate and more likely to be nonrecruited. This can be due to factors other than age and HF symptoms, which were, on average, not more severe than for other groups.

Gender did not affect the likelihood of participating. However, women represented less than one-third of the overall sample that entered the recruitment process and less than one-third of the participating group. This reflects difficulties in recruiting women with HF. Although the recruitment of women in HF studies has improved over time [50], the participation rate among women does still not reflect disease levels in the population [50,51]. In other studies, women were found to be more likely to decline the invitation to participate in digital health research [26,29]. This was not found for the sample in this study.

This study has some limitations. First, due to ethical clearance on collecting information about nonparticipants, only limited information on the patients was available. Such information did not include factors that can further explain participation in digital health research, such as level of education, digital skills, digital health literacy, and social participation. Future research should investigate the impact of such factors on both (non)recruitment and self-selection. On the other hand, it should be considered that collecting such detailed information, for example through a survey, might itself generate bias based on the nonresponse of some individuals. Therefore, although registered hospital information is limited, it can give an accurate description of selective non-participation. Second, for these analyses, reasons for refusal were only available grouped in main categories, which might have simplified the individual decision-making process for declining the invitation to participate in the HF-Wii study. Detailed reasons for declining the invitation to participate could have provided more insights into, for example, logistics-related and technology-related barriers to participation and allowed for a more elaborated description of the individual decision-making process. For example, in the full HF-Wii study it was found that 4% of the patients reported not wanting to participate because of the use of technology [32]. Future studies should further examine such reasons for refusal in combination with an analysis of the self-selection process. Third, as specific information on digital skills and technology-related aspects is not available, results from this study can also be relevant to the understanding of selective (non)participation in health research.

Selective nonparticipation in digital health research can prevent the production of results that appropriately inform research, policy, and practice on the impact of digital health technologies for the targeted populations. If participants and nonparticipants differ from one another, research results will not represent the target population of interest but rather a part of it. Groups that are often underrepresented in digital health research, such as people of older age and with poorer health, can be the most in need of accessing care and support [6] and can experience more barriers to using digital technologies [9,14]. Underrepresenting such groups implies overlooking their needs and interests, which are not necessarily expressed by their participating counterparts, and, as a result, miscalculating the impact of digital health technologies on the target population as a whole. Implementing digital health technologies that have been selectively evaluated might introduce further sources of exclusion and disadvantages to such groups with respect to their counterparts and contribute to widening old age inequalities.

We therefore recommend that a measure of selective nonparticipation is included in digital health research to identify overestimation and underestimation of the effects of digital health technologies due to the involvement of samples that do not reflect the target population. A measure of selective participation also allows researchers to employ further research strategies, such as focused recruitment of underrepresented groups or post-hoc adjustments of the results by weighting different groups depending on how well they are represented in the study sample.

Acknowledgments

This work was supported by the Swedish Research Council for Health, Working Life and Welfare (FORTE) (ref. 2014-4100). Moreover, the HF-Wii study from which this project collected data was supported by the Swedish National Science Council (K2013 ‐ 69X ‐ 22302 ‐ 01 ‐ 3), The Swedish Heart and Lung Association E085/12, The Swedish Heart ‐ Lung Foundation (20130340), the Vårdal Foundation (2014–0018), and the Medical Research Council of Southeast Sweden (FORSS 474681).

Authors' Contributions

AP contributed to the study conception, design, data collection planning, data analysis and coordination, and drafted the manuscript. SK, TJ, LK, and AS contributed to the study design, data collection planning, and data analysis. LK contributed to data collection. AMK devised the study and supervised its design, conduct, and coordination. All authors were involved in revising the manuscript critically for important intellectual content and have given their final approval of the version to be published.

Conflicts of Interest

None declared.

References

1. Czaja SJ. The Potential Role of Technology in Supporting Older Adults. Public Policy & Aging Report 2017;27(2):44-48 [FREE Full text] [doi: 10.1093/ppar/prx006]
2. Schulz R, Wahl H, Matthews JT, De Vito Dabbs A, Beach SR, Czaja SJ. Advancing the Aging and Technology Agenda in Gerontology. Gerontologist 2015 Oct;55(5):724-734 [FREE Full text] [doi: 10.1093/geront/gnu071] [Medline: 25165042]
3. Lindeman D. Improving the Independence of Older Adults through Technology: Directions for Public Policy. Public Policy & Aging Report 2017;27(2):49-52. [doi: 10.1093/ppar/prx011]
4. European Commission. 2018. Communication on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51628 [accessed 2020-05-22]
5. World Health Organisation. Global Observatory for eHealth. Geneva: World Health Organization: World Health Organization; 2016. Global diffusion of eHealth: Making universal health coverage achievable. Report of the third global survey on eHealth URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1 [accessed 2020-05-22]
6. Terraneo M. Inequities in health care utilization by people aged 50+: evidence from 12 European countries. Soc Sci Med 2015 Feb;126:154-163. [doi: 10.1016/j.socscimed.2014.12.028] [Medline: 25562311]
7. People in the EU - statistics on an ageing society.: Eurostat; 2017. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/People_in_the_EU_-_statistics_on_an_ageing_society [accessed 2020-05-22]
8. Lee C, Czaja S, Moxley J, Sharit J, Boot W, Charness N, et al. Attitudes Toward Computers Across Adulthood From 1994 to 2013. Gerontologist 2019 Jan 09;59(1):22-33 [FREE Full text] [doi: 10.1093/geront/gny081] [Medline: 29982458]
9. König R, Seifert A, Doh M. Internet use among older Europeans: an analysis based on SHARE data. Univ Access Inf Soc 2018 Jan 19;17(3):621-633. [doi: 10.1007/s10209-018-0609-5]
10. Helsper EJ, Reisdorf BC. The emergence of a “digital underclass” in Great Britain and Sweden: Changing reasons for digital exclusion. New Media & Society 2016 Mar 03;19(8):1253-1270. [doi: 10.1177/1461444816634676]
11. Olsson T, Samuelsson U, Viscovi D. At risk of exclusion? Degrees of ICT access and literacy among senior citizens. Information, Communication & Society 2017 Jul 20;22(1):55-72. [doi: 10.1080/1369118x.2017.1355007]
12. Fang M, Canham S, Battersby L, Sixsmith J, Wada M, Sixsmith A. Exploring Privilege in the Digital Divide: Implications for Theory, Policy, and Practice. Gerontologist 2019 Jan 09;59(1):e1-e15. [doi: 10.1093/geront/gny037] [Medline: 29750241]
13. Kim J, Lee HY, Christensen MC, Merighi JR. Technology Access and Use, and Their Associations With Social Engagement Among Older Adults: Do Women and Men Differ? J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):836-845. [doi: 10.1093/geronb/gbw123] [Medline: 28073816]
14. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
15. Vorrink SNW, Antonietti AMGEF, Kort HSM, Troosters T, Zanen P, Lammers JJ. Technology use by older adults in the Netherlands and its associations with demographics and health outcomes. Assist Technol 2017;29(4):188-196. [doi: 10.1080/10400435.2016.1219885] [Medline: 27548101]
16. Siren A, Knudsen SG. Older Adults and Emerging Digital Service Delivery: A Mixed Methods Study on Information and Communications Technology Use, Skills, and Attitudes. J Aging Soc Policy 2017;29(1):35-50. [doi: 10.1080/08959420.2016.1187036] [Medline: 27215167]
17. Wyatt S. Non-users also matter: The construction of non-users of the Internet. In: Oudshoorn N, Pinch T, editors. How Users Matter: The Co-construction of Users and Technology. Cambridge: MIT Press; 2003:67-79.
18. Merkel S, Kucharski A. Participatory Design in Gerontechnology: A Systematic Literature Review. Gerontologist 2019 Jan 09;59(1):e16-e25. [doi: 10.1093/geront/gny034] [Medline: 29788319]
19. Grates M, Heming A, Vukoman M, Schabsky P, Sorgalla J. New Perspectives on User Participation in Technology Design Processes: An Interdisciplinary Approach. Gerontologist 2019 Jan 09;59(1):45-57. [doi: 10.1093/geront/gny112] [Medline: 30260407]
20. Epstein S. Inclusion, Diversity,Biomedical Knowledge Making: The Multiple Politics of Representation. In: Oudshoorn N, Pinch T, editors. How Users Matter: The Co-Construction of Users and Technology. Cambridge: MIT Press; 2003:173-190.
21. Allemann H, Poli A. Designing and evaluating information and communication technology-based interventions? Be aware of the needs of older people. Eur J Cardiovasc Nurs 2020 Jan 23:1-3. [doi: 10.1177/1474515119897398] [Medline: 31969003]
22. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
23. Robinson L, Cotten SR, Ono H, Quan-Haase A, Mesch G, Chen W, et al. Digital inequalities and why they matter. Information, Communication & Society 2015 Mar 16;18(5):569-582. [doi: 10.1080/1369118X.2015.1012532]
24. DiMaggio P, Garip F. Network Effects and Social Inequality. Annu. Rev. Sociol 2012 Aug 11;38(1):93-118. [doi: 10.1146/annurev.soc.012809.102545]
25. Mair FS, Goldstein P, Shiels C, Roberts C, Angus R, O'Connor J, et al. Recruitment difficulties in a home telecare trial. J Telemed Telecare 2006;12 Suppl 1:26-28. [doi: 10.1258/135763306777978371] [Medline: 16884571]

26. Green BB, Anderson ML, Ralston JD, Catz S, Fishman PA, Cook AJ. Patient ability and willingness to participate in a web-based intervention to improve hypertension control. J Med Internet Res 2011 Jan 20;13(1):e1 [FREE Full text] [doi: 10.2196/jmir.1625] [Medline: 21371993]
27. Dodge HH, Katsumata Y, Zhu J, Mattek N, Bowman M, Gregor M, et al. Characteristics associated with willingness to participate in a randomized controlled behavioral clinical trial using home-based personal computers and a webcam. Trials 2014 Dec 23;15:508 [FREE Full text] [doi: 10.1186/1745-6215-15-508] [Medline: 25539637]
28. Foster A, Horspool KA, Edwards L, Thomas CL, Salisbury C, Montgomery AA, et al. Who does not participate in telehealth trials and why? A cross-sectional survey. Trials 2015 Jun 05;16(1):258 [FREE Full text] [doi: 10.1186/s13063-015-0773-3] [Medline: 26044763]
29. Broendum E, Ulrik CS, Gregersen T, Hansen EF, Green A, Ringbaek T. Barriers for recruitment of patients with chronic obstructive pulmonary disease to a controlled telemedicine trial. Health Informatics J 2018 Jun;24(2):216-224. [doi: 10.1177/1460458216667166] [Medline: 27638452]
30. Subramanian U, Hopp F, Lowery J, Woodbridge P, Smith D. Research in home-care telemedicine: challenges in patient recruitment. Telemed J E Health 2004;10(2):155-161. [doi: 10.1089/tmj.2004.10.155] [Medline: 15319045]
31. Jaarsma T, Klompstra L, Ben Gal T, Boyne J, Vellone E, Bäck M, et al. Increasing exercise capacity and quality of life of patients with heart failure through Wii gaming: the rationale, design and methodology of the HF-Wii study; a multicentre randomized controlled trial. Eur J Heart Fail 2015;17(7):743-748 [FREE Full text] [doi: 10.1002/ejhf.305] [Medline: 26139585]
32. Jaarsma T, Klompstra L, Ben Gal T, Ben Avraham B, Boyne J, Bäck M, et al. Effects of exergaming on exercise capacity in patients with heart failure: results of an international multicentre randomized controlled trial. Eur J Heart Fail 2020:1-11. [doi: 10.1002/ejhf.1754] [Medline: 32167657]
33. Klompstra L, Jaarsma T, Strömberg A. Exergaming to increase the exercise capacity and daily physical activity in heart failure patients: a pilot study. BMC Geriatr 2014 Nov 18;14(1):119 [FREE Full text] [doi: 10.1186/1471-2318-14-119] [Medline: 25407612]
34. Klompstra L, Jaarsma T, Strömberg A. An in-depth, longitudinal examination of the daily physical activity of a patient with heart failure using a Nintendo Wii at home: a case report. J Rehabil Med 2013 Jun;45(6):599-602 [FREE Full text] [doi: 10.2340/16501977-1151] [Medline: 23695704]
35. Ponikowski P, Voors A, Anker S, Bueno H, Cleland J, Coats A, ESC Scientific Document Group. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016 Jul 14;37(27):2129-2200 [FREE Full text] [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
36. Poli A, Kelfve S, Motel-Klingebiel A. A research tool for measuring non-participation of older people in research on digital health. BMC Public Health 2019 Nov 08;19(1):1487 [FREE Full text] [doi: 10.1186/s12889-019-7830-x] [Medline: 31703655]
37. Dolgin M, New York Heart Association. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. In: Dolgin M, editor. The Criteria Committee of the New York Heart Association. Boston: Little, Brown; 1994:1-334.
38. Williams R. Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects. The Stata Journal 2018 Nov 19;12(2):308-331. [doi: 10.1177/1536867x1201200209]
39. Mood C. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. European Sociological Review 2009 Mar 09;26(1):67-82. [doi: 10.1093/esr/jcp006]
40. Provencher V, Mortenson WB, Tanguay-Garneau L, Bélanger K, Dagenais M. Challenges and strategies pertaining to recruitment and retention of frail elderly in research studies: a systematic review. Arch Gerontol Geriatr 2014;59(1):18-24. [doi: 10.1016/j.archger.2014.03.006] [Medline: 24745811]
41. Taylor JS, DeMers SM, Vig EK, Borson S. The disappearing subject: exclusion of people with cognitive impairment and dementia from geriatrics research. J Am Geriatr Soc 2012;60(3):413-419. [doi: 10.1111/j.1532-5415.2011.03847.x] [Medline: 22288835]
42. van Heuvelen MJG, Hochstenbach JBM, Brouwer WH, de Greef MHG, Zijlstra GAR, van Jaarsveld E, et al. Differences between participants and non-participants in an RCT on physical activity and psychological interventions for older persons. Aging Clin Exp Res 2005 Jun;17(3):236-245. [doi: 10.1007/BF03324603] [Medline: 16110738]
43. Elskamp AB, Hartholt KA, Patka P, van Beeck EF, van der Cammen TJ. Why older people refuse to participate in falls prevention trials: a qualitative study. Exp Gerontol 2012;47(4):342-345. [doi: 10.1016/j.exger.2012.01.006] [Medline: 22310657]
44. Harrison JM, Jung M, Lennie TA, Moser DK, Smith DG, Dunbar SB, et al. Refusal to participate in heart failure studies: do age and gender matter? J Clin Nurs 2016 Apr 23;25(7-8):983-991 [FREE Full text] [doi: 10.1111/jocn.13135] [Medline: 26914834]
45. Chatfield MD, Brayne CE, Matthews FE. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. J Clin Epidemiol 2005 Jan;58(1):13-19. [doi: 10.1016/j.jclinepi.2004.05.006] [Medline: 15649666]

46. Thake M, Lowry A. A systematic review of trends in the selective exclusion of older participant from randomised clinical trials. Arch Gerontol Geriatr 2017;72:99-102. [doi: 10.1016/j.archger.2017.05.017] [Medline: 28618323]
47. Cherubini A, Oristrell J, Pla X, Ruggiero C, Ferretti R, Diestre G, et al. The persistent exclusion of older patients from ongoing clinical trials regarding heart failure. Arch Intern Med 2011;171(6):550-556. [doi: 10.1001/archinternmed.2011.31] [Medline: 21444844]
48. Guha K, McDonagh T. Heart failure epidemiology: European perspective. Curr Cardiol Rev 2013 May;9(2):123-127 [FREE Full text] [doi: 10.2174/1573403x11309020005] [Medline: 23597298]
49. Lazzarini V, Mentz RJ, Fiuzat M, Metra M, O'Connor CM. Heart failure in elderly patients: distinctive features and unresolved issues. Eur J Heart Fail 2013 Jul;15(7):717-723 [FREE Full text] [doi: 10.1093/eurjhf/hft028] [Medline: 23429975]
50. Gong IY, Tan NS, Ali SH, Lebovic G, Mamdani M, Goodman SG, et al. Temporal Trends of Women Enrollment in Major Cardiovascular Randomized Clinical Trials. Can J Cardiol 2019 May;35(5):653-660. [doi: 10.1016/j.cjca.2019.01.010] [Medline: 31030866]
51. Pressler SJ. Women With Heart Failure Are Disproportionately Studied as Compared With Prevalence. The Journal of Cardiovascular Nursing 2016;31(1):84-88. [doi: 10.1097/jcn.0000000000000212]

Abbreviations

AME: average marginal effect
FORTE: FORTE: Swedish Reserach Council for Health, Working Life and Welfare (Forskningsrådet för hälsa, arbetsliv och välfärd)
HF: heart failure
NYHA: New York Heart Association
OR: odds ratio
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 18.01.20; peer-reviewed by S Merkel, M Johansson; comments to author 26.02.20; revised version received 23.03.20; accepted 23.03.20; published 05.06.20.

Please cite as:
Poli A, Kelfve S, Klompstra L, Strömberg A, Jaarsma T, Motel-Klingebiel A
Prediction of (Non)Participation of Older People in Digital Health Research: Exergame Intervention Study
J Med Internet Res 2020;22(6):e17884
URL: http://www.jmir.org/2020/6/e17884/
doi:10.2196/17884
PMID:32501275

Original Paper

Physiological State and Learning Ability of Students in Normal and Virtual Reality Conditions: Complexity-Based Analysis

Mohammad H Babini[1], MSc; Vladimir V Kulish[2], PhD; Hamidreza Namazi[1], PhD

[1]School of Engineering, Monash University, Subang Jaya, Malaysia

[2]Faculty of Mechanical Engineering, Czech Technical University in Prague, Prague, Czech Republic

Corresponding Author:
Hamidreza Namazi, PhD
School of Engineering
Monash University
Jalan Lagoon Selatan
Subang Jaya, 47500
Malaysia
Phone: 60 172757619
Email: hamidreza.namazi@monash.edu

Abstract

Background: Education and learning are the most important goals of all universities. For this purpose, lecturers use various tools to grab the attention of students and improve their learning ability. Virtual reality refers to the subjective sensory experience of being immersed in a computer-mediated world, and has recently been implemented in learning environments.

Objective: The aim of this study was to analyze the effect of a virtual reality condition on students' learning ability and physiological state.

Methods: Students were shown 6 sets of videos (3 videos in a two-dimensional condition and 3 videos in a three-dimensional condition), and their learning ability was analyzed based on a subsequent questionnaire. In addition, we analyzed the reaction of the brain and facial muscles of the students during both the two-dimensional and three-dimensional viewing conditions and used fractal theory to investigate their attention to the videos.

Results: The learning ability of students was increased in the three-dimensional condition compared to that in the two-dimensional condition. In addition, analysis of physiological signals showed that students paid more attention to the three-dimensional videos.

Conclusions: A virtual reality condition has a greater effect on enhancing the learning ability of students. The analytical approach of this study can be further extended to evaluate other physiological signals of subjects in a virtual reality condition.

(J Med Internet Res 2020;22(6):e17945) doi:10.2196/17945

KEYWORDS

virtual reality; learning ability; brain; facial muscle; fractal theory

Introduction

Virtual reality refers to the subjective sensory experience of being immersed in a computer-mediated world. Accumulating evidence [1,2] points to the exciting opportunity and potential of integrating virtual reality technology in education environments, which can add elements of reality to improve understanding of complex subjects such as the life sciences (eg, biology and anatomy) compared to traditional classes in which students must imagine the structures for comprehension. In addition, students have been shown to pay more attention to the lecturer when making direct eye contact [3]. However, a lecturer is only able to look at one or two students at a time during a lecture. Therefore, presenting a virtual image of a lecturer to students might increase their attention and consequently improve their learning ability, which can be applied to an electronic learning environment.

Along with the growing empirical evidence that virtual reality is a valuable learning tool, further investigations are needed to study how the use of virtual reality can improve the learning ability of students. In addition, few studies have focused on changes in physiological signals to understand the effect of virtual reality on the human body. Previous studies in this field have compared the brain reaction of anxious participants at rest and under a virtual reality condition [4], compared electroencephalogram (EEG) signals in virtual reality and the traditional display condition [5], analyzed brain activity in response to increasing levels of task complexity in virtual reality [6], employed a deep-learning approach to improve the rate of excitement to well above the 90% accuracy level [7], and analyzed reactions of the heart and brain in different virtual reality environments [8].

The aim of the present study was to investigate the attention and learning ability of students using virtual reality technology. We also investigated the variability of students' physiological state (facial reaction) under the virtual reality condition. To study the attention and learning ability of students, we recorded their brain signals (ie, EEG signals) and to study facial reactions, we recorded their electromyography (EMG) signals. For comparison, we also recorded the brain signals and facial reactions of the students in a traditional classroom learning condition. EMG was used to capture the engagement of facial muscles during visual perception, which we expected to be more strongly affected under the virtual reality condition.

Since both EEG and EMG signals have complex patterns, we adopted complexity theory for our analysis. In other words, the concept of complexity was employed to define the structure of EEG and EMG signals. Complexity theory can help to characterize the behavior of a system with many parts that interact with each other in highly variable manners [9]. Specifically, we analyzed the recorded EEG and EMG signals using fractal theory, which can be used to quantify the complexity of a system (EEG and EMG signals in this case). Fractals are self-similar or self-affine objects that have complex structures [10]. A self-similar fractal has the same scaling exponent at every scale, whereas a self-affine fractal has different values of the scaling exponent at different scales. EEG and EMG signals are self-affine fractals that have a nonlinear structure. An object with a greater fractal dimension (as a measure of complexity) has a greater level of complexity [11]. Several studies have analyzed different types of physiological signals using fractal theory to date, including analyses of magnetoencephalography [12], galvanic skin response [13], heart rate [14], respiration [15], speech-evoked auditory brainstem response [16], eye movement [17], and human DNA [18]. Similarly, many studies have applied fractal analysis to investigate the nonlinear structure of EEG signals under different conditions, including the influence of auditory [19,20], olfactory [21], and visual [22,23] stimuli; brain diseases [24]; body movements [25,26]; and aging [27].

Some previous studies have also applied fractal theory to analyze EMG signals, including a decoded finger [28,29], hand [29-31], and functional movements and force patterns [29], along with analysis of the effect of complexity of walking on a path with respect to the leg muscle reaction [32]. However, to our knowledge, only one study has employed fractal theory to analyze the facial muscle reaction to date [33].

To analyze and compare the physiological conditions of subjects in virtual reality versus traditional class conditions, we used fractal analysis to relate the complexity of EEG and EMG signals to the nature of the viewed videos.

Methods

Study Design

We aimed to analyze students' physiological state and learning ability under the three-dimensional (3D) virtual reality condition in comparison to those recorded under the traditional two-dimensional (2D) condition. For analysis of the physiological state, we chose EEG and EMG signals as indicators of the brain and muscle response, respectively. EMG signals were selected for the facial muscle reaction since the subjects were stimulated using visual stimuli. For this purpose, we used fractal theory to analyze the complexity of facial EMG and EEG signals. The fractal dimension, as the main quantitative measure of fractal theory, indicates the complexity of the process in which greater values of a fractal dimension reflect greater complexity of the object.

Various methods have been developed to calculate the fractal dimension, which are mainly based on the entropy concept. In this study, we used the box-counting method to calculate the fractal dimension [34]. In the box-counting algorithm, the object of interest is covered with boxes of the same size (ε). The number of boxes (*N*) required to cover the object is then counted. This process is repeated several times, while the box size keeps changing in each step. Finally, the slope of the regression line fitted to a log-log plot of the number of boxes versus the scale is calculated as an estimate of the fractal dimension for the object under consideration [35]:

Equation (2) defines the so-called generalized fractal dimension of order *c* [35]:

where is the Rényi entropy of order *c*, and the probability of occurrence (r_j) is defined as:

In Equation (3), the total time of the signal value occurrence within the jth value interval is denoted by t_j, whereas *T* represents the total duration of the recorded signal [36].

In this experiment, we showed the students 6 sets of videos (3 videos in the 2D condition and the same 3 videos in the 3D condition), and then investigated the reaction of the brain and facial muscle under both the 2D and 3D conditions using fractal theory to assess the students' attention to the videos.

In addition, to investigate the learning ability, we designed three questions based on the content of each video (9 questions in total for the 3 videos) that were asked to the students after watching each video in each condition. This questionnaire allowed for assessing the extent to which the students retained and learned the content of the videos.

Data Collection and Analysis

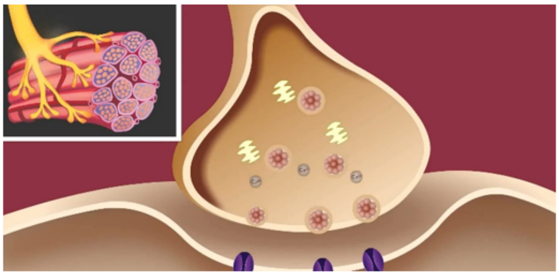
All procedures from recruiting subjects to conducting the experiment were approved by the Monash University Human Research Ethics Committee (MUHREC; approval number 20965). The study was carried out in accordance with the approved guidelines.

We conducted the experiment with 9 healthy students from Monash University Malaysia. We explained the experiment to the participants and then asked them several questions about their health conditions. Since mental disorders, some medications, as well as drinking beverages that contain alcohol or caffeine affect brain activity and cause inconsistent results, we excluded potential participants within these categories. In addition, participants were excluded if they had consumed beverages containing alcohol or caffeine within 24 hours before the experiments. The students who were deemed to be suitable for experiments signed the consent forms and were included in the study.

We conducted the experiment in a quiet room to isolate the participants from other external stimuli that could potentially affect the recorded EEG and EMG signals. The participants were asked to sit comfortably on a chair during the experiment, and were instructed to focus on watching the videos without engaging in any other task.

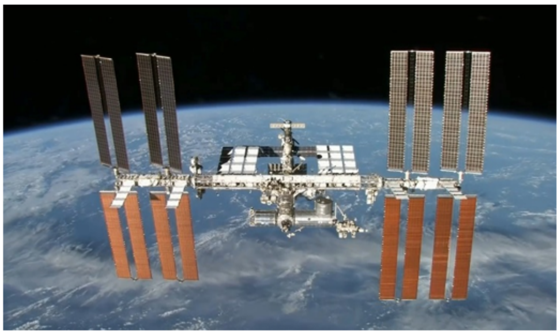
As mentioned above, we chose 6 sets of videos (3 videos in the 2D condition and the same 3 videos in the 3D condition) for our experiment. The 2D videos were selected from YouTube, which were then converted into 3D videos for our experiment. The first and second videos (same content) were about biology (Multimedia Appendix 1 and Multimedia Appendix 2), the third and fourth videos (Multimedia Appendix 3 and Multimedia Appendix 4) were about architecture, and the fifth and sixth videos (Multimedia Appendix 5 and Multimedia Appendix 6) were about space. Some screenshots from these videos are shown in Figure 1.

Figure 1. Representative screenshots of scenes from three different videos.



(a) Biology

(b) Architecture



(c) Space

The videos were displayed to the participants via a mobile phone. The participants watched the 2D videos with the naked eye, whereas the 3D videos were viewed through VeeR MINI VR Glasses (VeeR, Atlanta, GA, USA) (see Figure 2) in front of the mobile phone. We noninvasively recorded EEG and facial EMG signals from the participants using an EMOTIV EPOC+

14 Channel Mobile EEG headset (Emotiv, San Francisco, CA, USA) and Shimmer EMG device (Shimmer, Ireland) with a sampling frequency of 128 Hz and 256 Hz, respectively. As shown in Figure 2, the EEG device was placed on the participant's head and five electrodes of the EMG device were connected to the facial muscles.

Figure 2. Data collection from a participant.

First, we recorded EEG and EMG signals from the participants for 2 minutes while they watched the first 2D video. When the video was complete, we then asked the participants three questions related to the content. After a 1-minute rest period, the participants watched the first video again in the 3D condition for 2 minutes. The content of this video was identical to that of the 2D video, except that it was presented in 3D mode. Three questions were then asked about the content of the 3D video, followed by another period of rest for 1 minute. We continued this procedure to collect EEG and EMG signals from the participants (along with the responses to content-related questions) with the third, fourth, fifth, and sixth videos (each video lasted for 2 minutes), providing the participants with 1 minute of rest between watching the videos. The data collection was repeated for each participant in the second session to validate the repeatability of results.

Initially, we preprocessed the raw data to remove noise. For this purpose, we wrote a set of codes in MATLAB (MathWorks, Natick, MA, USA) based on the Butterworth filter. The frequency bands of 1-40 Hz and 25-125 Hz were chosen for filtering the EEG and EMG signals, respectively. Of note, two electrodes of the EEG device had some disconnection problems during data collection; therefore, we processed the collected data only from the other 12 electrodes.

After initial filtering, we proceeded with the analysis by computing the fractal dimension of the recorded EEG and EMG signals. The computation of the fractal dimension was based on the box-counting algorithm using boxes with sizes (1/2, 1/4, 1/8, etc) as the scaling factor. Although we recorded 120 seconds of data during each period of watching the 2D and 3D videos, we analyzed only 118.2 seconds of each dataset. This selection was due to the fact that the devices did not always have a consistent sampling frequency, which caused the data recording to be less than 2 minutes long, leading to a difference of a few seconds in the duration of collected data among some participants.

After confirming the normal distribution of the data, statistical analysis of the computed fractal dimension for EEG and EMG signals was performed to assess the effect of stimulation on variations of the fractal dimension of EEG and EMG signals using one-way repeated-measures analysis of variance (ANOVA). We also conducted the Student *t* test to compare the difference in mean values of EEG or EMG signals between the 2D and 3D condition. *P*<.05 was considered to reflect a statistically significant difference in our analysis.

Results

The variations of fractal dimensions of EEG signals for the first to the sixth visual stimuli are shown in Table 1. As mentioned above, the first, third, and fifth stimuli refer to the 2D condition, whereas the second, fourth, and sixth stimuli refer to the 3D condition.

Table 1. Fractal dimension of EEG[a] signals with the first to sixth stimuli.

Stimulus	Fractal dimension of EEG signal
First	1.7027
Second	1.7266
Third	1.7196
Fourth	1.7222
Fifth	1.6928
Sixth	1.7272

[a] EEG: electroencephalogram.

Based on the result of ANOVA (*F*=7.6334, *P*<.001), the effect of stimulation (2D and 3D) on variations of fractal dimensions of the EEG signal was significant. As shown in Table 1, for all stimuli, the EEG signal recorded from the participants in the 3D condition had a greater fractal dimension compared to that recorded in the 2D condition. Since the fractal dimension reflects the complexity of the signal, this result indicated that the EEG signal is more complex in response to 3D visual stimuli compared to 2D visual stimuli. In other words, the human brain becomes more engaged with a stimulus when it is presented in the 3D condition compared to the 2D condition. Differences between the mean values of the EEG signal from the first and second stimuli (*P*=.001) and from the fifth and sixth stimuli (*P*<.001) were greater than the difference between the third and fourth stimuli (*P*=.72). This suggested that the participants' brains were more engaged with the second and sixth stimuli compared to the third stimuli. This result is reasonable given that the second and sixth stimuli mainly contained animated scenes, whereas the fourth stimulus included more photos with less animated scenes. Therefore, the difference between the fractal dimension of the EEG signals in the third and fourth stimuli was lower than that observed under the other conditions.

The variations of the fractal dimension of the EMG signals for the first to the sixth visual stimuli are summarized in Table 2.

Table 2. Fractal dimension of EMG[a] signals with the first to sixth stimuli.

Stimulus	Fractal dimension of EMG signal
First	1.2361
Second	1.2594
Third	1.2554
Fourth	1.2580
Fifth	1.2488
Sixth	1.2675

[a] EMG: electromyography.

Based on the result of ANOVA (*F*=0.2468, *P*=.94), the effect of stimulation (2D and 3D) on variations of the fractal dimension of the EMG signals was not significant. Upon receiving a visual stimulus (2D or 3D), the brain processes the stimulus and then sends the impulses to the facial muscles. Therefore, the stimulus should have a greater effect on the brain than on the facial muscles, which explains why there was a significant effect of the stimuli on variations of EEG signals but not on the facial muscles.

As shown in Table 2, for all stimuli, the EMG signal had a greater value of the fractal dimension in response to 3D videos compared to 2D videos, indicating that the EMG signal is more complex in response to 3D videos compared to 2D videos. In other words, the facial muscles are more engaged with stimuli where they are presented in 3D rather than in 2D. In addition, the difference between the mean values of the EMG signal in the first and second stimuli (*P*=.43) and the fifth and sixth stimuli (*P*=.56) was greater than that between the third and fourth stimuli (*P*=.93). This indicates that the participants' facial muscles were more engaged with the second and sixth stimuli compared to the third stimuli. As mentioned above, this difference can be explained by the content of the videos, in which the second and sixth stimuli contained more animated scenes compared to the fourth video.

Despite these differences among videos, there was no significant difference in the fractal dimension of the EMG signal between each pair of stimuli. This suggests that although presenting the videos in 3D caused some changes in the muscle reaction, these changes were not substantial. Comparison of the results for EMG and EEG signals indicates that changing the visual stimulus from 2D to 3D could cause significant variations in the complexity of the EEG signal, but not in the EMG signal. Therefore, changes in the state of the brain are greater when changing a visual stimulus from 2D to 3D.

Moreover, evaluating the relationship between variations of EEG and EMG signals can provide further insight. The brain controls all parts of the human body, including the reactions of the facial muscle. When exposed to 2D or 3D videos as visual stimuli, the brain sends impulses to the facial muscles. Therefore, when the brain is more engaged with the stimuli, the muscle reaction will also be greater, which is reflected in the greater variations in the fractal dimension of the EMG signal.

The rate of correct responses to the questions posed after watching the 3D video was 92.60%, which was higher than that obtained after the 2D videos at 80.87%. This difference suggested that the 3D videos resulted in greater attention paid to the details of videos and therefore increased the learning ability of the students.

Discussion

In this study, we compared the effect of virtual reality on students' learning ability and physiological state with those recorded in a normal 2D condition based on watching 3 sets of videos each presented in 2D and 3D. We simultaneously recorded EEG and facial EMG signals of the participants during stimulation. Overall, the EEG and EMG signals had greater fractal dimensions in the 3D video condition, indicating that both the brain and facial muscles have a greater reaction to 3D videos compared to 2D videos. In addition, videos with more animated scenes resulted in a significantly greater brain reaction compared with that resulting from watching a video with less animated scenes, as reflected by the lack of a significant difference in the fractal dimension of EEG signals between 2D and 3D conditions. For the EMG analysis, although the 3D condition caused greater reaction in the facial muscle, there was no significant difference from the reaction recorded under the 2D condition.

We also examined the learning ability of the students after watching each video by asking them several content-related questions, demonstrating improved learning ability after watching 3D videos than 2D videos (92.60% vs 80.87% correct answers). These results clearly showed that students pay more attention to videos when they are presented in 3D. The present study offers a step forward compared to previous studies that only analyzed the learning ability or brain reaction [4-7] in a virtual reality condition without considering the reaction of facial muscles and investigating how that reaction correlates with brain activity.

The method of analysis employed in this study can be extended to investigate other physiological signals of students in a virtual reality condition. For instance, we can analyze how the heart rate changes in a 3D condition compared to a 2D condition. We can also expand this work by applying other types of stimuli, including olfactory stimuli, while students are watching videos in the virtual reality condition and investigate the effect of these additional stimuli on their learning ability. Developing a model between input (videos) and outputs (human physiological signals) is another important aspect of future work in this regard. For this purpose, we can benefit from different tools such as machine learning [37-39] and fractional-based mathematical equations [40]. Such analysis could allow for predicting human conditions (physiological signals) before exposure to different stimuli, providing guidance on the types of videos and characteristics of videos that are most likely to arouse the attention of students and facilitate learning. These efforts therefore have great importance in advancing research on students' learning ability and can provide strong recommendations to educational institutions.

Acknowledgments

The authors are very grateful to the Faculty of Engineering at Monash University for providing a travel grant and the Department of Thermodynamics and Fluid Mechanics at Czech Technical University in Prague for providing financial support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
2D-biology.
[MP4 File (MP4 Video), 45931 KB - jmir_v22i6e17945_app1.mp4]

Multimedia Appendix 2
3D-biology.
[MP4 File (MP4 Video), 63761 KB - jmir_v22i6e17945_app2.mp4]

Multimedia Appendix 3
2D Architecture.

[MP4 File (MP4 Video), 13047 KB - jmir_v22i6e17945_app3.mp4]

Multimedia Appendix 4
3D Architecture.
[MP4 File (MP4 Video), 35200 KB - jmir_v22i6e17945_app4.mp4]

Multimedia Appendix 5
2D Space.
[MP4 File (MP4 Video), 29040 KB - jmir_v22i6e17945_app5.mp4]

Multimedia Appendix 6
3D Space.
[MP4 File (MP4 Video), 36793 KB - jmir_v22i6e17945_app6.mp4]

References

1. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
2. Markowitz DM, Laha R, Perone BP, Pea RD, Bailenson JN. Immersive Virtual Reality Field Trips Facilitate Learning About Climate Change. Front Psychol 2018 Nov 30;9:2364. [doi: 10.3389/fpsyg.2018.02364] [Medline: 30555387]
3. Yin Y. Contact with My Teacher's Eyes. Phenemol Practice 2013 Jul 11;7(1):69-81. [doi: 10.29173/pandpr20104]
4. Tarrant J, Viczko J, Cope H. Virtual Reality for Anxiety Reduction Demonstrated by Quantitative EEG: A Pilot Study. Front Psychol 2018 Jul 24;9:1280. [doi: 10.3389/fpsyg.2018.01280] [Medline: 30087642]
5. Tauscher JP, Wolf Schottky F, Grogorick S, Maximilian Bittner P, Mustafa M, Magnor M. Immersive EEGvaluating Electroencephalography in Virtual Reality. 2019 Aug 15 Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces (VR); 2019; Osaka, Japan p. 1794-1800. [doi: 10.1109/VR.2019.8797858]
6. Dey A, Chatburn A, Billinghurst M. Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality. 2019 Aug 15 Presented at: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); 2019; Osaka, Japan p. 220-226. [doi: 10.1109/vr.2019.8797840]
7. Teo J, Chia JT. EEG-based excitement detection in immersive environments: An improved deep learning approach. 2018 Apr 10 Presented at: 3rd International Conference on Applied Science and Technology (ICAST'18); 2018; Penang, Malaysia. [doi: 10.1063/1.5055547]
8. Marín-Morales J, Higuera-Trujillo JL, Greco A, Guixeres J, Llinares C, Scilingo EP, et al. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. Sci Rep 2018 Sep 12;8(1):13657. [doi: 10.1038/s41598-018-32063-4] [Medline: 30209261]
9. Johnson S. Emergence: The Connected Lives of Ants, Brains, Cities, and Software. New York: Scribner; 2012.
10. Kiew CL, Brahmananda A, Islam KT, Lee HN, Vernier SA, Saraar A, et al. Complexity-based analysis of the relation between tool wear and machine vibration in turning operation. Fractals 2020 Feb 03;28(01):2050018. [doi: 10.1142/s0218348x20500188]
11. Kiew C, Brahmananda A, Islam KT, Lee HN, Venier SA, Saraar A, et al. Analysis of the relation between fractal structures of machined surface and machine vibration signal in turning operation. Fractals 2020 Feb 07;28(01):2050019. [doi: 10.1142/s0218348x2050019x]
12. Namazi H, Jafari S. Decoding of wrist movements' direction by fractal analysis of magnetoencephalography (MEG) signal. Fractals 2019 May 30;27(02):1950001. [doi: 10.1142/s0218348x19500014]
13. Omam S, Babini MH, Sim S, Tee R, Nathan V, Namazi H. Complexity-based decoding of brain-skin relation in response to olfactory stimuli. Comput Methods Programs Biomed 2020 Feb;184:105293. [doi: 10.1016/j.cmpb.2019.105293] [Medline: 31887618]
14. Tapanainen JM, Thomsen PEB, Køber L, Torp-Pedersen C, Mäkikallio TH, Still A, et al. Fractal analysis of heart rate variability and mortality after an acute myocardial infarction. Am J Cardiol 2002 Aug 15;90(4):347-352. [doi: 10.1016/s0002-9149(02)02488-8] [Medline: 12161220]
15. Namazi RH. Fractal-based analysis of the influence of music on human respiration. Fractals 2017 Nov 21;25(06):1750059. [doi: 10.1142/s0218348x17500591]
16. Mozaffarilegha M, Namazi H, Tahaei AA, Jafari S. Complexity-Based Analysis of the Difference Between Normal Subjects and Subjects with Stuttering in Speech Evoked Auditory Brainstem Response. J Med Biol Eng 2018 Jun 1;39(4):490-497. [doi: 10.1007/s40846-018-0430-x]
17. Alipour H, Towhidkhah F, Jafari S, Menon A, Namazi H. Complexity-Based Analysis of the Relation Between Fractal Visual Stimuli and Fractal Eye Movements. Fluct Noise Lett 2019 Jul 16;18(03):1950012. [doi: 10.1142/s0219477519500123]

18. Namazi H, Akrami A, Hussaini J, Silva ON, Wong A, Kulish VV. The fractal based analysis of human face and DNA variations during aging. Biosci Trends 2017 Jan 16;10(6):477-481. [doi: 10.5582/bst.2016.01182] [Medline: 27795483]
19. Alipour ZM, Khosrowabadi R, Namazi H. Fractal-based analysis of the influence of variations of rhythmic patterns of music on human brain response. Fractals 2018 Oct 30;26(05):1850080. [doi: 10.1142/s0218348x18500809]
20. Namazi H, Khosrowabadi R, Hussaini J, Habibi S, Farid AA, Kulish VV. Analysis of the influence of memory content of auditory stimuli on the memory content of EEG signal. Oncotarget 2016 Aug 30;7(35):56120-56128 [FREE Full text] [doi: 10.18632/oncotarget.11234] [Medline: 27528219]
21. Kawano K. Electroencephalography and Its Fractal Analysis During Olfactory Stimuli. In: Olfaction and Taste XI. 1994 Presented at: Proceedings of the 11th International Symposium on Olfaction and Taste and of the 27th Japanese Symposium on Taste and Smell; July 12-16, 1993; Kaikan, Sapporo, Japan p. 668-672. [doi: 10.1007/978-4-431-68355-1_269]
22. Namazi H, Ala TS, Bakardjian H. Decoding of steady-state visual evoked potentials by fractal analysis of the electroencephalographic (EEG) signal. Fractals 2019 Jan 15;26(06):1850092. [doi: 10.1142/s0218348x18500925]
23. Ahmadi-Pajouh MA, Ala TS, Zamanian F, Namazi H, Jafari S. Fractal-based classification of human brain response to living and non-living visual stimuli. Fractals 2018 Oct 30;26(05):1850069. [doi: 10.1142/s0218348x1850069x]
24. Namazi H, Aghasian E, Ala TS. Fractal-based classification of electroencephalography (EEG) signals in healthy adolescents and adolescents with symptoms of schizophrenia. Technol Health Care 2019;27(3):233-241. [doi: 10.3233/THC-181497] [Medline: 30829625]
25. Namazi H, Ala TS, Kulish V. Decoding of upper limb movement by fractal analysis of electroencephalogram (EEG) signal. Fractals 2018 Oct 30;26(05):1850081. [doi: 10.1142/s0218348x18500810]
26. Kamal SM, Sim S, Tee R, Nathan V, Aghasian E, Namazi H. Decoding of the relationship between human brain activity and walking paths. Technol Health Care 2019 Nov 14:Epub ahead of print. [doi: 10.3233/THC-191965] [Medline: 31796717]
27. Namazi H, Jafari S. Age-based variations of fractal structure of EEG signal in patients with epilepsy. Fractals 2018 Sep 25;26(04):1850051. [doi: 10.1142/s0218348x18500512]
28. Namazi H. Fractal-based classification of electromyography (EMG) signal in response to basic movements of the fingers. Fractals 2019 May 30;27(03):1950037. [doi: 10.1142/s0218348x19500373]
29. Namazi H. Fractal-based classification of electromyography (EMG) signal between fingers and hand's basic movements, functional movements, and force patterns. Fractals 2019 Jul 11;27(04):1950050. [doi: 10.1142/s0218348x19500506]
30. Namazi H. Decoding of hand gestures by fractal analysis of electromyography (EMG) signal. Fractals 2019 May 30;27(03):1950022. [doi: 10.1142/s0218348x19500221]
31. Namazi H, Jafari S. Decoding of simple hand movements by fractal analysis of electromyography (EMG) signal. Fractals 2019 Jul 11;27(04):1950042. [doi: 10.1142/s0218348x19500427]
32. Kamal SM, Sim S, Tee R, Nathan V, Namazi H. Complexity-Based Analysis of the Relation between Human Muscle Reaction and Walking Path. Fluct Noise Lett 2020 Jan 28:2050025. [doi: 10.1142/s021947752050025x]
33. Arjunan SP, Kumar DK. Measuring complexity in different muscles during sustained contraction using fractal properties of SEMG signal. 2018 Oct 29 Presented at: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018; Honolulu, HI p. 5656-5659. [doi: 10.1109/embc.2018.8513544]
34. Li J, Du Q, Sun C. An improved box-counting method for image fractal dimension estimation. Pattern Recognition 2009 Nov;42(11):2460-2469. [doi: 10.1016/j.patcog.2009.03.001]
35. Qadri MO, Namazi H. Fractal-Based Analysis of the Relation Between Surface Finish and Machine Vibration in Milling Operation. Fluct Noise Lett 2019 Jun 28;19(01):2050006. [doi: 10.1142/s0219477520500066]
36. Kulish V, Sourin A, Sourina O. Human electroencephalograms seen as fractal time series: mathematical analysis and visualization. Comput Biol Med 2006 Mar;36(3):291-302. [doi: 10.1016/j.compbiomed.2004.12.003] [Medline: 16446162]
37. Rezvani S, Wang X, Pourpanah F. Intuitionistic Fuzzy Twin Support Vector Machines. IEEE Trans Fuzzy Syst 2019 Nov;27(11):2140-2151. [doi: 10.1109/tfuzz.2019.2893863]
38. Pourpanah F, Lim CP, Hao Q. A reinforced fuzzy ARTMAP model for data classification. Int J Mach Learn Cyber 2018 Jun 15;10(7):1643-1655. [doi: 10.1007/s13042-018-0843-4]
39. Pourpanah F, Zhang B, Ma R, Hao Q. Anomaly Detection and Condition Monitoring of UAV Motors and Propellers. 2018 Dec 27 Presented at: 2018 IEEE Sensors; October 28-31, 2018; New Delhi, India p. 1-4. [doi: 10.1109/ICSENS.2018.8589572]
40. Namazi H, Kulish VV. Fractional Diffusion Based Modelling and Prediction of Human Brain Response to External Stimuli. Comput Math Methods Med 2015;2015:148534. [doi: 10.1155/2015/148534] [Medline: 26089955]

Abbreviations

2D: two-dimensional
3D: three-dimensional
ANOVA: analysis of variance
EEG: electroencephalogram
EMG: electromyography

Edited by G Eysenbach; submitted 04.02.20; peer-reviewed by F Pourpanah, S Jafari; comments to author 25.02.20; revised version received 04.03.20; accepted 14.03.20; published 01.06.20.

Please cite as:
Babini MH, Kulish VV, Namazi H
Physiological State and Learning Ability of Students in Normal and Virtual Reality Conditions: Complexity-Based Analysis
J Med Internet Res 2020;22(6):e17945
URL: https://www.jmir.org/2020/6/e17945
doi:10.2196/17945
PMID:32478661

Original Paper

Simulator Pre-Screening of Underprepared Drivers Prior to Licensing On-Road Examination: Clustering of Virtual Driving Test Time Series Data

David Grethlein[1,2*], BSc; Flaura Koplin Winston[1,3,4*], MD, PhD; Elizabeth Walshe[3,5*], PhD; Sean Tanner[1,6*], BA, MA; Venk Kandadai[1*], MPH; Santiago Ontañón[2*], PhD

[1]Diagnostic Driving, Inc, Philadelphia, PA, United States
[2]Computer Science Department, Drexel University, Philadelphia, PA, United States
[3]Center for Injury Research and Prevention, Children's Hospital of Philadelphia, Philadelphia, PA, United States
[4]Perelmen School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[5]Annenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA, United States
[6]Geography Department, Rutgers University, New Brunswick, NJ, United States
[*]all authors contributed equally

Corresponding Author:
David Grethlein, BSc
Diagnostic Driving, Inc
705 S 50th Street
Philadelphia, PA, 19143
United States
Phone: 1 215 315 3204
Email: david@diagnosticdriving.com

Abstract

Background: A large Midwestern state commissioned a virtual driving test (VDT) to assess driving skills preparedness before the on-road examination (ORE). Since July 2017, a pilot deployment of the VDT in state licensing centers (VDT pilot) has collected both VDT and ORE data from new license applicants with the aim of creating a scoring algorithm that could predict those who were underprepared.

Objective: Leveraging data collected from the VDT pilot, this study aimed to develop and conduct an initial evaluation of a novel machine learning (ML)–based classifier using limited domain knowledge and minimal feature engineering to reliably predict applicant pass/fail on the ORE. Such methods, if proven useful, could be applicable to the classification of other time series data collected within medical and other settings.

Methods: We analyzed an initial dataset that comprised 4308 drivers who completed both the VDT and the ORE, in which 1096 (25.4%) drivers went on to fail the ORE. We studied 2 different approaches to constructing feature sets to use as input to ML algorithms: the standard method of reducing the time series data to a set of manually defined variables that summarize driving behavior and a novel approach using time series clustering. We then fed these representations into different ML algorithms to compare their ability to predict a driver's ORE outcome (pass/fail).

Results: The new method using time series clustering performed similarly compared with the standard method in terms of overall accuracy for predicting pass or fail outcome (76.1% vs 76.2%) and area under the curve (0.656 vs 0.682). However, the time series clustering slightly outperformed the standard method in differentially predicting failure on the ORE. The novel clustering method yielded a risk ratio for failure of 3.07 (95% CI 2.75-3.43), whereas the standard variables method yielded a risk ratio for failure of 2.68 (95% CI 2.41-2.99). In addition, the time series clustering method with logistic regression produced the lowest ratio of false alarms (those who were predicted to fail but went on to pass the ORE; 27.2%).

Conclusions: Our results provide initial evidence that the clustering method is useful for feature construction in classification tasks involving time series data when resources are limited to create multiple, domain-relevant variables.

(J Med Internet Res 2020;22(6):e13995) doi:10.2196/13995

KEYWORDS

simulated driving assessment; on-road exam; machine learning; adolescent; child; support vector machines; humans; accidents, traffic; cause of death; licensure; automobile driving; motor vehicle; motor vehicles

Introduction

Background

According to the Centers for Disease Control and Prevention, motor vehicle crashes (MVC) are the leading cause of death in adolescents aged 12 to 19 years in the United States [1]. MVC risk is disproportionately high among novice drivers, particularly those aged 16 to 17 years, and peaks immediately after licensure during the first months of unsupervised driving [2,3]. Previous studies have demonstrated that the majority of these crashes early in licensure are attributed to critical driving errors because of skill deficits, inexperience, and inattention/distraction rather than recklessness and deliberate risk taking [4-7]. Research has demonstrated initial evidence for the ability of web-based screening tests (eg, web-based assessment of cognitive impairments) to predict poor driving simulator performance [8]. However, there remains a critical need for a screening test to quantify skill level at the time of licensure to identify those who are underprepared to drive safely.

In 2017, a large Midwestern state’s driver licensing agency aimed to address the need to identify underprepared license applicants (likely to fail the on-road examination [ORE]) by developing and deploying a new portable virtual driving test (VDT) as a potential prescreen tool for ORE [9]. The VDT was built on the Ready-Assess platform [10] (Diagnostic Driving, Inc) and was designed as a safe, reliable, and portable method for evaluating novice drivers’ preparedness to respond to common and serious potential crash scenarios [11-13]. Ready-Assess incorporates and expands on the scenarios and metrics of the previously validated simulated driving assessment (SDA) [14] to provide a variety of traffic situations to manage that go beyond what is feasible to assess during an ORE.

Objectives

The VDT assessment (which lasts approximately 7 min) includes both high-risk and common driving scenarios. The potential crash scenarios included in the VDT driving environments were determined by the National Highway Traffic Safety Administration [15,16], including intersections (3- and 4-way, stop sign, and traffic light), curved roads, rear-end events (lead car brakes suddenly), and hazard zones (construction zones, school zones, and pedestrian crosswalks). These typical driving scenarios were specified by subject matter experts from the state's licensing and motor vehicle safety body encompass a variety of settings (eg, urban and rural) and on-road elements (eg, school buses, ambulances, pedestrians, and hazards) to reproduce local driving environments. As a result, the VDT provides an opportunity for a wide variety of driving responses.

Consequently, a large quantity of applicant time series data was generated from the VDT assessment. To efficiently create a prediction (ORE pass/fail) scoring algorithm, steps must be taken to reduce the dimensionality of the data for analysis. Many traditional machine learning (ML) methods (*out-of-the-box*) require multiple domain-specific features that are well defined and created before being evaluated in an ML pipeline. Given the task of predicting the ORE outcome at the individual level, manually defining features with optimized predictive capabilities is an extremely resource-intensive task and often reliant on subject matter expertise (SME) and considerable domain knowledge.

Therefore, by leveraging data collected from the VDT pilot, this study aimed to develop and conduct an initial evaluation of a novel ML-based classifier using limited domain knowledge and minimal feature engineering to reliably predict applicant pass/fail on the ORE. Such methods, if proven useful, could be applicable to the classification of other time series data collected within the medical field and other settings.

Methods

Apparatus

The VDT used in this pilot was delivered by Diagnostic Driving, Inc [17] in collaboration with Children’s Hospital of Philadelphia (CHOP) and designated state-employed highly experienced driving examiners (SMEs). The VDT software evolved from a prior laboratory-based SDA into a highly scalable, portable, and self-directed tool (Figure 1) that could be delivered on ubiquitous hardware without the need for additional personnel to help facilitate its delivery (eg, research assistants and administrative personnel; Multimedia Appendices 1 and 2).

Figure 1. Workstation Setup: 1) Standard Monitor, 2) Standard Desktop Computer, 3) Off-the-shelf USB Steering Wheel and Pedals.

Virtual Driving Test Implementation Procedure

Applicants were directed to an available VDT workstation in the testing facility by state personnel, and applicants were asked to put on headphones (Multimedia Appendix 2) and enter the provided unique identification number that could be linked to their ORE results (Figure 2).

The VDT is a self-directed workflow (containing both on-screen and voice-directed instructions) that typically takes less than 15 min to complete after the applicant logs in to the VDT workstation with a unique identification number provided by state personnel. A short orientation video introduces the importance of safe driving and provides the applicant with an overview of the stages of the VDT workflow. This is followed by a simulated *practice drive*, a short orientation drive with built-in instructions to orient the applicant to all VDT controls (eg, turn signals, transmission controls, and navigation system) and to allow the applicant to test drive basic maneuvers (eg, steering, accelerating, braking, and 90° turns) on a course without additional traffic. This drive ends with a brief comprehension test that evaluates the driver applicant's ability to manage the controls and to follow basic instructions covered in it.

Next, the applicant begins the simulated *assessment drive*, a planned route through a randomly assigned environment selected from a bank of 10 possible environments (eg, *city 1*, *city 2*, ..., *city 10*), all of which contain variations of common and serious crash scenarios. In all the environments, driver applicants are never explicitly prompted to react to changing traffic conditions or traffic controllers (eg, *grant pedestrians in the crosswalk right of way* and *wait for the light to turn green*). They are, however, given navigational instructions to follow a planned route (eg, *turn left at the stop sign* and *shift into the right lane*). The assessment drive has no explicit time limit and finishes once the applicant brings his or her simulated vehicle to the end of the planned route.

On completion of the assessment drive, the applicant receives a series of 3 debriefing questions to answer on the screen (using a 5-point Likert scale: 1=strongly disagree and 5=strongly agree) to assess his or her ability to understand VDT directions and general comfort with the VDT controls.

Figure 2. Virtual driving test workstation.

Dataset

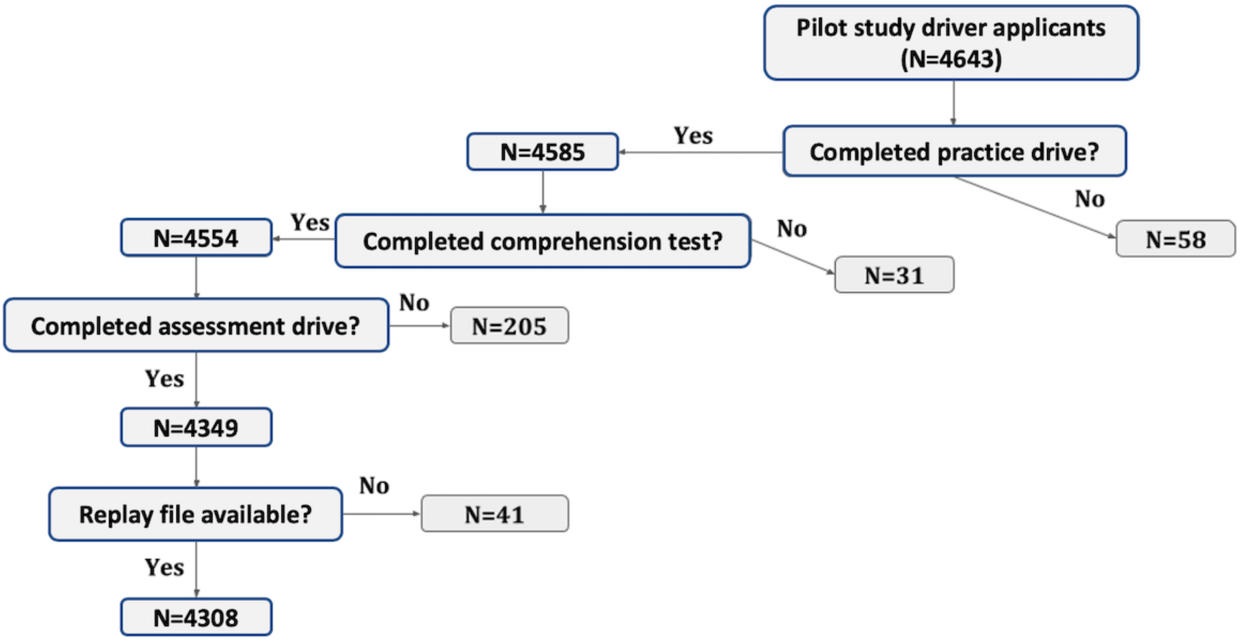
We received a deidentified dataset containing individually linked VDT and ORE performance data for 4643 driver applicants from 3 licensing facilities collected between July 2017 and March 2018. These data were collected by a large Midwestern state during a pilot phase of implementing the VDT in collaboration with Diagnostic Driving, Inc, and CHOP to test its utility as a screening tool in the context of a high-demand and established state government licensing workflow. During this pilot phase, no additional research or demographic information (eg, age and sex) was collected, and all driver applicants who completed the VDT went on to complete the ORE regardless of VDT performance.

In addition, neither the driver applicant nor the licensing examiner had access to the VDT results; therefore, no bias because of VDT performance was introduced when implementing the ORE. Driver applicant ORE outcomes were linked to VDT data and entered into a deidentified dataset that was shared with the research team. CHOP’s institutional review board determined that this study does not meet the criteria for human subjects research primarily because (1) data were not obtained by the researchers through intervention or interaction with the individual and (2) no identifiable private information was included in the dataset received.

Derivation of Sample

Figure 3 shows the derivation of the sample included in the analyzable dataset. Of the 4643 driver applicants who attempted the VDT in the pilot, the vast majority completed it: 58 (1.2%) applicants did not complete the practice drive, 31 (0.7%) did not complete the comprehension test, and 205 (4.4%) did not complete the assessment drive (descriptions of comprehension and assessment drives are given in the Virtual Driving Test Implementation Procedure section). Driver applicants who did not complete the entire VDT workflow, 6.3% (294/4643) were excluded from our analysis. SMEs from the licensing facilities provided information on why some applicants could not complete the workflow, and reasons included applicant had a language barrier, applicant did not understand the instructions, applicant was frustrated with the VDT software, applicant was called for their ORE earlier than expected, applicant walked away from the VDT workstation, applicant experienced symptoms of simulator sickness, and applicant elected not to continue (<1%). In addition to not completing the entire VDT workflow, an additional 41 (0.9%) applicants who completed the workflow did not have their replay files (raw time series information) successfully uploaded to the VDT cloud server (because of internet connectivity issues). These cases were also removed from our analysis, and the final analyzable sample included 4308 (92.8%) driver applicants.

Figure 3. Sample derivation: data from 335 (7.2%) enrolled driver applicants were excluded from the final sample of 4308 because they either did not complete the VDT workflow or their assessment replay file was unavailable for analysis. VDT: virtual driving test.



Data Collected: Time Series Channels

The VDT software is fully cloud based and collects raw time series–driving performance data, sampled at 60 Hz, including the driver's vehicle position; driver inputs (ie, steering wheel, brake, and throttle); additional driver attributes (ie, velocity); and information regarding ambient vehicles, pedestrians, and environmental objects. This raw and rich representation is then downsampled (for the purpose of storage) at approximately 10 Hz with the true elapsed time between contiguous frames recorded, with the added benefit of reducing the replay file size for permanent cloud storage while retaining the faithfulness of the time series representation. As each *replay file* (the downsampled recording of a VDT assessment) is stored as a multichannel sequence of frames averaging about 8 min in length downsampled to roughly 10 Hz, it provides more than 30,000 data points of information regarding driving performance.

Each frame of the recorded time series data includes several raw values regarding the applicant's physical interaction with the VDT workstation, including (1) the percentage of complete brake depression (with 1=full depression/stop) and throttle pedals (with 1=full depression/maximal acceleration), (2) the signed percentage of steering wheel rotations both left (negative) and right (positive) of its resting position, and (3) the position and heading of the applicant's vehicle within the simulated environment. Also recorded in each frame is the use of turn signals and steering wheel buttons that allow the driver to scan left or right to look for oncoming traffic.

The *lane offset* (sometimes referred to as lane position) [18] of the applicant's vehicle within its lane on the simulated road is computed for each frame and recorded as one of the time series channels. These channels comprise each applicant's VDT assessment recording and are used as the basis for feature set construction to represent driving performance.

Outcome Variable

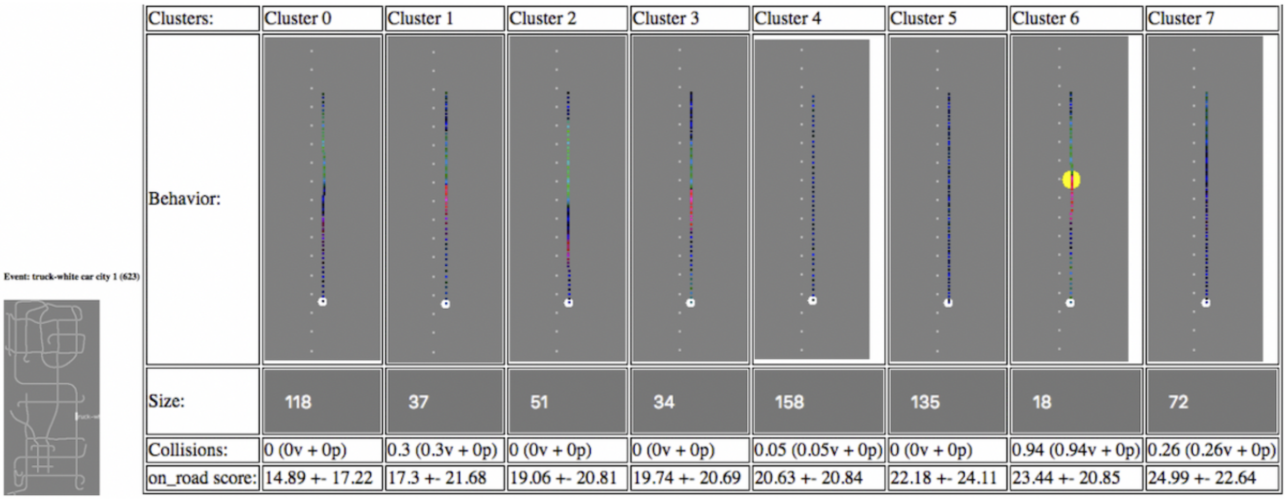
For all driver applicants who completed the VDT, the result of the ORE is provided in a numeric form (score), where a higher score indicates a more severe accumulation of infractions cited by the examiner during their ORE. According to the state licensing agency's ORE scoring protocol, any driver with a score greater than or equal to 26 fails the ORE and any driver with a score less than 26 passes the ORE (a score of 0 indicates a perfect score). These data (both numerical and dichotomous pass/fail representations) for each applicant are linked to the corresponding VDT record and added to the deidentified dataset. For the purposes of internal analysis, we define the *gray zone* of ORE scores between 20 and 35, where drivers either barely pass or accumulate just enough infractions to fail the ORE.

Analytical Procedure

To maximize our ability to reliably predict an applicant's likelihood of failing the ORE, we evaluated 2 alternative ways to represent the time series data. The first, standard approach (which we call the *variables* feature set) involves the reduction of the driver applicant's input time series data into manually created variables that represent driver behavior These variables further represent both continuous features (eg, velocity) and count/dichotomous data (eg, crash). The second method is novel and involves viewing the environments as a series of predefined *event zones* to which the applicant is exposed (eg, a crosswalk, an intersection, or a school zone). The corresponding applicant's input time series data (how the applicant responds to each of these event zones) are then partitioned according to these event zones. The range of applicant behaviors (combinations of steering, brake, and throttle on the simulated roadway) is then clustered for each segment without any preprocessing or manual reduction of the data. This *time series clustering* feature set has no shared features with the *variables* feature set.

To build classifiers of driver performance, we constructed these 2 feature sets as input for supervised ML algorithms (Figure 4).

Figure 4. Time Series Clustering for highlighted event zone with cluster centers representing prototypical driving behaviors in that zone.



Clusters:	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Behavior:								
Size:	118	37	51	34	158	135	18	72
Collisions:	0 (0v + 0p)	0.3 (0.3v + 0p)	0 (0v + 0p)	0 (0v + 0p)	0.05 (0.05v + 0p)	0 (0v + 0p)	0.94 (0.94v + 0p)	0.26 (0.26v + 0p)
on_road score:	14.89 +- 17.22	17.3 +- 21.68	19.06 +- 20.81	19.74 +- 20.69	20.63 +- 20.84	22.18 +- 24.11	23.44 +- 20.85	24.99 +- 22.64

Standard Method for Data Reduction: Variables Feature Set

For every simulated drive, there are events corresponding to common and serious potential crash scenarios. Applicants' responses that have been tracked and tabulated in real time from within the VDT-simulated environment (eg, reduction of speed entering a construction zone) are augmented with manually created variables from the stored data, which indicate aberrant or hazardous driving behaviors (eg, number of instances applicant ran a red traffic light).

A summary of speed management within a sample is represented by the global maximum, mean, and median of the applicant vehicle velocities as well as the ratio of the applicant vehicle's velocity to the posted speed limit recorded. The same statistics on velocity and speed ratio are computed for the following event zones: crosswalks, school zones, construction zones, playground zones, and banking curves. To flush out adherence to the posted speed limit, the global percentages of each sample driven more than 10, 15, and 20 mph above and below the posted speed limit are used as potential predictors.

These values, along with other computed driving performance metrics, result in a representation of each sample as a vector of 67 engineered features plus the *outcome* variable (ORE pass/fail). With this representation of the tabulated data, supervised learning methods are deployed using standard 10-fold cross-validation; we examined the resulting confusion matrices. To evaluate the predictive power of this representation, we examined the resulting confusion matrices.

Novel Method for Automated Data Reduction: Time Series Clustering Feature Set

Each applicant's stored time series data were truncated to begin at the first downsampled frame with a simulated speed above 0.5 m per second to align replay files via a common starting condition. As assessment drives are of nontrivial length (on average, approximately 8 min in duration), relying on global measurements to compare driving behaviors may wash out highly predictive features from smaller segments of the assessment's planned route.

The *time series clustering* representation attempts to derive new features from sections of the drives that are designed to elicit specialized behaviors in response to environmental or traffic conditions. Unlike the standard variables approach, this new approach derives features automatically without the need for manual data reduction. Time series samples are grouped by the simulated environment to which the applicant is assigned, where each environment comprises a sequence of stationary event zones defined by *entry* and *exitwaypoints*. Given a sample that drives through a particular event zone, a *subinterval* of the time series is extracted that corresponds to the vehicle traveling between the zone's *entry* and *exit* waypoints. This subinterval is grouped in a set with all other subintervals from the drives that pass through the designated event zone.

The purpose of clustering is to reveal *prototypical behaviors* so that drivers can be categorized within each event zone (as defined by their lane position, throttle, brake, accelerator, and steering wheel usage in each frame). For example, at a crosswalk, there might be a cluster for those who slow to a stop and another for those who speed through the crosswalk without slowing. The variance in behaviors should be represented in the cluster centers, where all sample subintervals assigned to a given cluster have been determined to be most similar to the cluster center's *prototypical behavior*.

Speed and traffic conditions determine each driver's subinterval length in an event zone. Thus, to compare subintervals of unequal lengths, dynamic time warping [19,20] is used as the (dis)similarity measure, where the time steps of pairs of subintervals are aligned to minimize the time-warped aggregate difference between them. *k*-Medoids [21], a derivative of *k-means*, is used to pigeonhole the n samples into k clusters (*k*=8 in all the reported experiments) by electing k initial subintervals to act as *medoids*, cluster centers representing prototypical behaviors. A dissimilarity matrix is constructed, where each cell compares subintervals χ_i and numerically. The subinterval χ_i is assigned to cluster , with the most similar medoid at round *r* of iteration. The algorithm elects new medoids in each round by identifying the candidate subinterval within each cluster that maximizes the aggregate similarity (ie, minimizes dissimilarity) between it and all other cluster members.

At every round of clustering the *inertia*, the sum of squared residual dissimilarities between subintervals and their assigned medoids is computed. If the newly elected medoids reduce inertia from the round before, the cycle repeats itself, although we imposed a limitation of 100 repetitions. Once *k-Medoids* converge on a clustering that cannot be improved upon, the dissimilarities between each sample subinterval χ_i to medoids are arranged as a feature vector.

This method of reporting the medoid dissimilarities can be thought of as *soft clustering*, where we report the relative *membership* of each subinterval to all clusters. As the initial medoids are randomly selected, 20 initializations of clustering are run for each event zone, using the medoids that minimize the overall inertia to represent the range of prototypical behaviors for the given event zone. This is done to select the most *compact* clusters where samples grouped together minimize aggregate dissimilarity from the medoids, distilling the prototypical behaviors that best represent the range of behaviors observed.

Illustrated on the far-left side of Multimedia Appendix 3 is a bird's-eye view of 1 of the 10 simulated environments' track, with a specific event zone highlighted (eg, *prescripted traffic interaction and avoid collision response to white car ahead suddenly stops in roadway*). Depicted on the right side of Multimedia Appendix 3 is the range of prototypical behaviors exhibited by the identified medoids while navigating through the indicated event zone.Beneath the depictions of each prototypical behavior is recorded the number of subintervals in its cluster, the percentage of those subintervals that had collisions in the defined event zone (both with vehicles and pedestrians), as well as the means and standard deviations of ORE scores from all cluster members.

With 166 different event zones in our dataset, each sample was converted to a concatenated feature vector of 1328 medoid dissimilarity features ($k \times 166$). For samples that never encounter a given event zone, the dissimilarities to the generated medoids are left as *k* blank entries. With the data table of medoid dissimilarities constructed as features and every sample having an *actual* pass/fail *outcome label*, supervised learning is trained and tested on the dataset using standard leave-one-out 10-fold cross-validation.

Supervised Learning Methods

Supervised learning methods evaluated features (both observed and derived) to establish their ability to predict a given driver applicant's *outcome label* (ORE pass/fail). Specifically, all methods were evaluated using a standard 10-fold cross-validation in an attempt to minimize the overfitting of any model. We evaluated 2 classification methods within the Waikato Environment for Knowledge Analysis [22,23]: logistic regression and support vector machines (SVMs).

Logistic Regression

Logistic regression ML algorithms classify a given sample based on computing its probability of belonging to one of the binary classes defined by the pass/fail *outcome label*. Given the predicted probability that a sample has either passed or failed, a decision rule is used to classify the sample:

In this example, the outcome label is based on actual ORE pass/fail, and the probability threshold used in the decision rule is derived from 0.0 to 1.0, with a step size of 0.01, to explore the threshold parameter's effect on performance metrics.

Support Vector Machines

SVMs attempt to compute the optimal decision surface for partitioning datasets along with binary class values (in our case, *pass* and *fail*) by maximizing the margin of separation between samples known to be members of different classes [24]. The core assumption of an SVM model is that there exists a transformed space in which the dataset may be linearly partitioned via a hyperplane. As it is computationally costly to transform every sample, a *kernel function* was used to relate pairs of samples in the transformed space.

For all results using SVM classification, we elected to use the *radial basis function* (RBF) as our kernel because our initial experimentation suggested that our feature values resemble Gaussian distributions. A coarse grid search was used to narrow the ranges of parameters considered for building an effective classifier [25]. SVMs for binary classification are primarily configured using 2 different values: a *cost* parameter (which determines the degree of influence of data points far from the decision surface in the algorithm) and a *gamma* parameter (which controls the variance of the Gaussian functions that make up the RBF kernel). We dyadically iterated the *cost* from 2^{-5} to 2^{-15} by doubling. To perform a grid search in a 2-dimensional parameter space, we also dyadically iterated the *gamma* parameter from 2^{-10} to 2^{6} by doubling.

Evaluation Metrics

As previously described, the primary goal of the VDT pilot was to reliably assess applicant preparedness before taking the ORE, thereby maximizing the safety of driving examiners and providing an opportunity for prepared applicants to take the ORE. The evaluation metrics described in the following sections were used to evaluate a given model's predictive ability while addressing the goals of the VDT pilot.

Confusion Matrix and Evaluation Metrics

To predict a binary outcome of the ORE with confidence, a confusion matrix (see Table 1) for each classification model was generated to evaluate the predictive ability of our classifiers.

Our evaluation metrics included accuracy, algorithm fail rate, false alarm rate, ratio of false alarm, and risk ratio.

Accuracy was defined as the ratio of the total number of correctly classified cases to the total sample population:

Algorithm fail rate was defined as the percentage of total cases the classifier predicted as fail:

False alarm rate was defined as the percentage of total cases the classifier misclassified as fail:

The *ratio of false alarms* was defined as the ratio of cases of misclassified *fail* to the total cases where the classifier predicted *fail*:

The *risk ratio* was defined as the relative risk (RR) of failing the ORE when the classifier predicted *fail* versus predicting *pass*. The 95% CIs were also calculated for risk ratios:

SMEs within the licensing agency recommended that a satisfactory classifier would consist of a model that maximized both accuracy and risk ratio while minimizing the false alarm rate (applicants who failed the VDT but went on to pass their ORE).

Table 1. A confusion matrix, showing the 4 quadrants: FF (fail-fail), FP (fail-pass), PF (pass-fail), and PP (pass-pass).

Confusion matrix	Fail ORE[a]	Pass ORE
Fail VDT[b]	FF	FP
Pass VDT	PF	PP

[a] ORE: on-road examination.

[b] VDT: virtual driving test.

Fitness of Models

In addition to the evaluation metrics previously described, we also evaluated model fitness using *receiver operator characteristic* (ROC) curves to compare our predictions against a random binary classification. An ROC curve illustrates a model's performance using different parameterizations. Specifically, it plots the model's *true positive rate* (TPR) to its *false positive rate* (FPR). For our data, TPR (also known as sensitivity) is the ratio of driver applicants who are correctly predicted to fail the ORE to the overall population of applicants who went on to fail the ORE:

FPR is the ratio of driver applicants incorrectly predicted to fail the ORE to the overall population of applicants who went on to pass the ORE:

Classifiers that have equal TPR and FPR do not perform well, as there is less certainty that the classification of any particular sample is reliable. ROC curves that deviate significantly from the line TPR=FPR represent classifiers that perform better at distinguishing the 2 classes. A perfect classifier would have a TPR of 1.0 and an FPR of 0.0, where a random binary classifier would have these 2 values be the same. A common metric used to summarize the relationship between TPR and FPR over a range of model parameterizations is the *area under the curve* (AUC). AUC is the numerical integral of the ROC curve (FPR is the independent variable and TPR is the dependent variable), where a higher percentage of the total area coverable (TPR of 1.0 for all values of FPR) should indicate models that are more fit to reliably predict a class value (eg, ORE outcome) [26].

Results

Classifier Results

We summarize the classifier results with the evaluation metrics previously described in Table 2. In this table, results from 4 classifiers were reported: feature sets generated from 2 methods (*variables* and the novel approach) and *time series clustering*, used in 2 ML methods (logistic regression and SVM). Overall, the results from the *time series clustering* approach provided similar results to the *variables* approach. More specifically, logistic regression using the time series clustering approach to generate features produced the best classifier based on our evaluation metrics (highest accuracy: 76.2%; highest risk ratio: 3.07; 95% CI 2.75-3.43). In addition, this classifier minimizes the ratio of false alarms compared with the others (27.2%).

Table 2. Summary results of the evaluation metrics obtained from 4 classifiers: variables+logistic regression, variables+support vector machine, time series clustering+logistic regression, and series time clustering+support vector machine.

Classifier results	Standard method (variables)		Novel method (time series clustering)	
	Logistic regression	SVM[a]	Logistic regression	SVM
Accuracy, %	76.1	75.4	76.2	74.9
Fail rate, %	6.6	3.4	3.5	3.6
False alarm rate, %	2.6	1.3	1.0	1.7
Ratio of false alarms, %	38.5	37.7	27.2	45.9
Relative risk (95% CI)	2.684 (2.409-2.991)	2.581 (2.250-2.961)	3.071 (2.747-3.434)	2.223 (1.906-2.592)
True positive rate, %	15.9	8.3	10.0	7.8
False positive rate, %	3.4	1.7	1.3	2.2

[a] SVM: support vector machine.

Logistic Regression

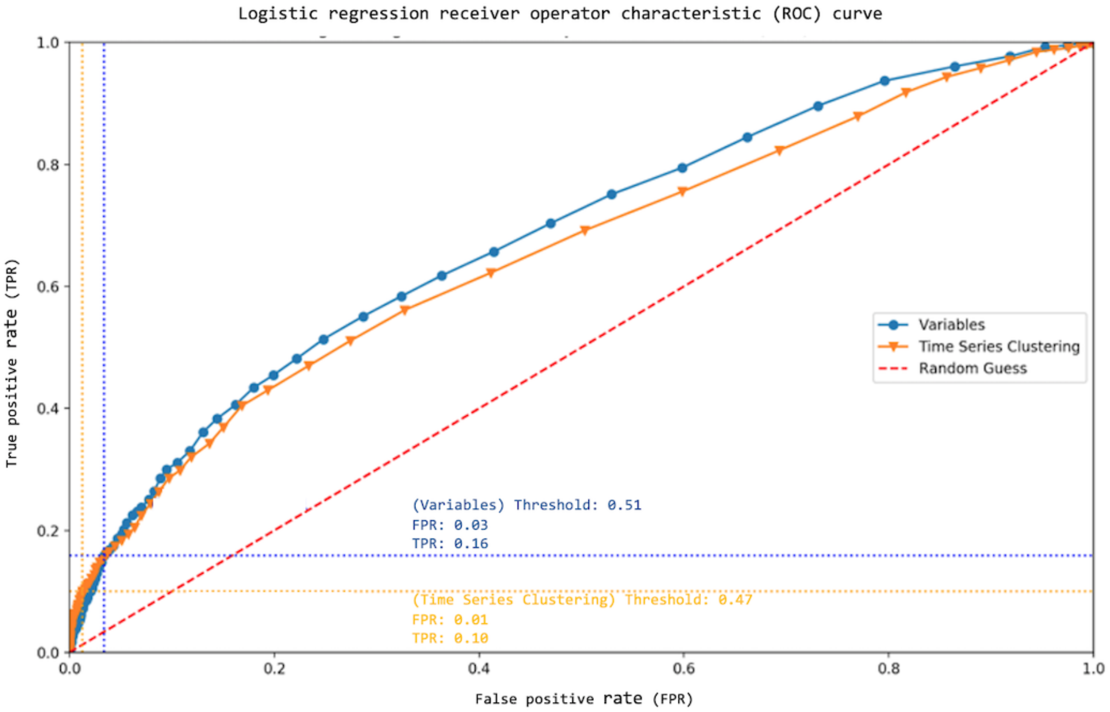
Figure 5 highlights the results for logistic regression with the highest observed classification accuracies for both representations (*variables* and *times series clustering*). The most promising parameterization of logistic regression using the *variables* representation as input for the VDT has an FPR of 3.4% and a TPR of 15.9%, whereas the most promising parameterization of logistic regression using the *time series clustering* representation as input for the VDT has an FPR of 1.3% and a TPR of 10.0%. Moreover, we observed that the *time series clustering* representation yielded a VDT classifier with a smaller ratio of false alarms (27.2%) than the most accurate VDT classifier using the *variables* representation (38.5%).

The threshold value yielding the model with the highest classification accuracy is highlighted in Figure 5. The highlighted parameterizations (yielding the highest classification accuracy, marked by the vertical dotted lines in the figure) demonstrate VDTs that show predictive promise, although there is room for improvement. Using logistic regression with different threshold values, the AUC for the *variables* feature set's ROC curve is 0.682, whereas the AUC for the *time series clustering* feature set's ROC curve is 0.656.

As seen in Figure 6 of Multimedia Appendix 4, logistic regression using the *variables* representation of the dataset produces a classifier that fails 6.9% of those taking the VDT, where 39% of those predicted to fail the ORE actually went on to pass (ratio of false alarms). The false alarms represent only 3% of all drivers, with 37% of the false alarms having an ORE score between 20 and 25, meaning these drivers nearly failed the ORE. Within the group of drivers misclassified (24% of all driver applicants), 30% of samples have an ORE score in the *gray zone* (20-35, with a score of ≤25 resulting in a pass). With this model, drivers failed by the VDT are 2.7 times more likely to fail the ORE (RR=2.68; 95% CI 2.41-2.97).

Using logistic regression, the *time series clustering* representation of the dataset produces a classifier that fails 4% of those taking the VDT, with a ratio of false alarms of 27%. Only 1% of all drivers are false alarms, where 25% of false alarms nearly fail the ORE with scores between 20 and 25. For all misclassified drivers (24% of all driver applicants), 25% of the samples had an ORE score in the *gray zone*. Using this model, drivers failed by the VDT are more than 3 times more likely to fail the ORE (RR=3.07; 95% CI 2.75-3.43). Overall, the *time series clustering* representation used as input for the logistic regression algorithm produces a more lax classifier that fails approximately half as many drivers as using the *variables* representation as input. However, when predicting failure on the ORE, logistic regression using the *time series clustering* representation is more often correct than with the *variables* representation (72% correct predictions of *fail* versus 62%).

Figure 5. Receiver operator characteristic curves for logistic regression using the variables and time series clustering feature sets for iterated logistic cutoff threshold values. Points in the bottom left represent models with the lowest thresholds (more people pass the virtual driving test), whereas points in the top right represent models with the highest thresholds (fewer people pass the virtual driving test).



Support Vector Machines

As Figure 7 in Multimedia Appendix 5 shows, the most accurate parameterization of SVM using the *variables* representation as input for the VDT has an FPR of 1.7% and a TPR of 8.2%, whereas the most promising parameterization of SVM using the *time series clustering* representation as input for the VDT has an FPR of 2.2% and a TPR of 7.8%. Moreover, we observed that the *time series clustering* representation yielded a VDT with a larger ratio of false alarms (45%) than the most accurate VDT using the *variables* representation (38%).

Loose grid search of SVM’s 2 parameters reinforces the notion that the 2 classes of ORE pass/fail are highly comingled within both feature spaces. Figure 8 in Multimedia Appendix 6 contains surface plots for both feature sets showing the interpolated landscapes of risk ratio and accuracy (vertical axis) over the SVM gamma and C (horizontal/lateral axes) parameters sampled in bivariate grid search.

The SVM parameterization with the highest observed classification accuracy using the *variables* representation of the dataset produces a classifier that fails 3.5% of those taking the VDT, with 38% of those predicted to fail the ORE being false alarms. The false alarms comprise 1.4% of all drivers, with 33% of the false alarms having nearly failed the ORE, where scores between 20 and 25 indicate nearly failing the ORE. Of the driver applicants misclassified (24% of all driver applicants), 26% of samples had an ORE score in the *gray zone* (20-35, with a score of ≤25 resulting in a pass). With this model, drivers failed by the VDT are 2.6 times more likely to fail the ORE, corresponding to the plateau of risk ratios illustrated in Figure 8 of Multimedia Appendix 6.

Using the *time series clustering* feature set, the SVM parameterization yielding the highest classification accuracy failed 3.6% of people taking the VDT, where 45% of those predicted to fail the ORE are false alarms. With 1.6% of all driver applicants being a false alarm, 13% of the false alarms nearly failed the ORE with scores between 20 and 25. This VDT misclassifies 25% of all driver applicants, where 26% of these drivers have an ORE score in the *gray zone*. With this model, drivers failed by the VDT are 2.3 times more likely to fail the ORE.

Discussion

Principal Findings

This study specifically aimed to evaluate a novel method for automatically classifying user-generated time series data versus a traditional and resource-intensive approach (classification by manually creating features [variables] using domain knowledge and SME). When compared with the standard method of feature selection (using manually created variables), the clustering method demonstrated a similar to higher accuracy in predicting ORE results. More specifically, it outperformed the standard method by accurately identifying the most underprepared applicants (those applicants who failed the ORE).

Although time series clustering is not a completely new idea, our particular application using a two-stage analysis of critical subintervals (our event zones) is novel. Moreover, our application does not suffer from the known problem of *sliding-window partitioningthen cluster* methods, which have been shown to be inconsistent in producing meaningful predictions [27]. In this study, we leveraged contextual similarities in scripted scenarios written to elicit specialized driving behavior in response to changes in the environment along the route, including ambient traffic [28]. Clustering the behaviors exhibited in these event zones leads to highly correlated clusters [29] that are quantifiably useful for the task of predicting ORE pass/fail.

The clustering is useful beyond simple binary prediction: Knowing how the data break down into groups allowed us to formulate a profile made of *prototypical behavior*, which represents the traits exhibited by the cluster members [30]. Only a small portion of the sample (433 of the 4308 drivers) had an ORE score greater than 45 (scores <26 result in passing). As we dealt with a heavily weighted distribution of ORE pass/fail (roughly 3 to 1), we carefully scrutinized the composition of the clusters formed; we also focused on identifying dangerous drivers (higher accumulation of critical driving errors) instead of optimizing overall predictive accuracy for ORE pass/fail.

Although the clusters did not cleanly separate all those who failed from those who passed, the features generated and then classified were able to identify outliers in the data rather easily. In many cases, these outliers presented aberrant driving behaviors, such as collisions with vehicles, collisions with pedestrians, and driving off the road. As it is meant to be a prescreening test, the purpose of VDT was never to identify all those who would go on to fail the ORE but rather to isolate the driver applicants who present a danger to themselves and others on the road. Operationally, the VDT should not restrict eligible driver applicants from taking the ORE; with that in mind, our novel approach shows promise in confidently classifying truly unprepared driver applicants. As a result, we considered evaluation metrics such as risk ratio instead of accuracy to keep state ORE examiners out of harm's way.

Limitations

This study had some limitations. First, the *variables* approach that we used did not represent the entire domain set of VDT performance metrics that could be derived and used. Many of the initial variables focused heavily on global measures (eg, measures across the entire simulated drive) rather than focusing on event-specific measures (eg, performance on curves and intersections). Owing to this, our analyses may not have taken into account the full variation in differential VDT performance (as we would expect that event-specific measures would be able to further explain this variability).

Second, the sample size available precluded model formulation specific to the simulated environment (eg, 10 separate models). Although assignment to clusters was performed for each of the 10 unique environments, only 1 dataset determined model formulation: each driver applicant record included a series of cluster assignments and their corresponding ORE score. Future research with larger samples will account for the different simulated environments and evaluate them individually.

Finally, because of restrictions placed on implementing the VDT in a busy licensing workflow, our limited dataset did not include demographic information and confounding variables collected for this sample. Age and sex are known risk factors for crashes among novice drivers [31] and may explain some variability in performances among new drivers seeking a driver's license.

Conclusions and Future Directions

This study provided initial evidence that the *time series clustering* feature set when used as input produced classifiers that performed just as well and, in some cases, better than those using the traditional *variables* approach. Future work will evaluate this method with larger sample sizes and potentially integrate it with other known methods to develop an optimized classifier. We also plan to explore more sophisticated variants of this approach, specifically expanding on a predefined number of clusters and using approaches such as the *silhouette method* [32] to determine an appropriate number of clusters for grouping behaviors exhibited by driver applicants in each event zone. Alternatively, we may attempt to maximize feature variance between clusters using a secondary *analysis of variance* clustering refinement procedure [33]. For the purpose of improved differential diagnostic capabilities in time series classification tasks, we intend to pursue several avenues of research that demonstrate potential for producing distinct prototypical user behaviors.

Acknowledgments

The authors would like to thank the individuals involved in the VDT pilot for collecting these data and providing our team with a deidentified dataset.

Conflicts of Interest

All authors had no contact with the driver license applicants and were not directly involved with data collection in the VDT pilot program. FW and VK disclose their intellectual property and financial interest in Diagnostic Driving, Inc and CHOP's institutional interest in Diagnostic Driving, Inc. Diagnostic Driving, Inc created the VDT. FW serves as the Chief Scientific Advisor, and VK serves as the Chief Executive Officer of Diagnostic Driving, Inc. FW operates under a Conflict Management Plan from CHOP and the University of Pennsylvania (Penn). EW also works at CHOP and Penn with FW. SO operates under a Conflict Management Plan from Drexel University. DG and ST are both nonshareholding employees of Diagnostic Driving, Inc.

Multimedia Appendix 1

Workstation Setup: 1) Standard Dell Monitor, 2) Standard Dell PC, 3) Logitech G29 USB Steering Wheel and Pedals.
[DOCX File , 111 KB - jmir_v22i6e13995_app1.docx]

Multimedia Appendix 2
Virtual driving test workstation.
[DOCX File , 965 KB - jmir_v22i6e13995_app2.docx]

Multimedia Appendix 3
Time Series Clustering for highlighted event zone with cluster centers representing prototypical driving behaviors in that zone. In the eight sub-plots, “prototypical behaviors” represent the derived cluster medoids for a specific event zone. Plots with more densely congregated and colorized pixels indicate the driver was moving more slowly through the given event zone than a plot with sparser, colorized pixels. The white dashes indicate the road median; time series subintervals begin at the white circle and the position of the driver applicant in each frame is a colorized pixel. Red corresponds to usage of the brakes, green corresponds to the throttle, light blue represents steering, and the spacing of these pixels are determined by the vehicle’s speed in the given zone. Collisions with other vehicle or pedestrians are indicated by red circles, collisions with static environmental objects are indicated by yellow circles. Displayed underneath are the means and standard deviations of actual ORE scores of samples in each cluster. In addition, displayed is the number of samples in each cluster and the average number of collisions with vehicles and pedestrians in a given cluster.
[DOCX File , 728 KB - jmir_v22i6e13995_app3.docx]

Multimedia Appendix 4
Performance Metrics and Confidence Intervals for Logistic Regression using “Variables” and “Time Series Clustering” representations as feature sets for iterated logistic cut-off threshold values.
[DOCX File , 180 KB - jmir_v22i6e13995_app4.docx]

Multimedia Appendix 5
The most successful SVM parameterizations for both feature sets learn to distinguish data which is largely homogeneous with high “Cost” penalties for misclassification errors. The optimal trade-off observed between TPR and FPR occurs with a fail rate of approximately 3.5%, about 40% of samples failed are false alarms.
[DOCX File , 221 KB - jmir_v22i6e13995_app5.docx]

Multimedia Appendix 6
We consistently observed more SVM parameterizations that better predicted ORE pass/fail using the “Variables” feature set as input than for “Time Series Clustering”. Zero is imputed for undefined Risk Ratios in these plots, such results originate from VDTs with parameterizations that fail no one.
[DOCX File , 342 KB - jmir_v22i6e13995_app6.docx]

References

1. Miniño AM. Centers for Disease Control and Prevention. 2010. Mortality Among Teenagers Aged 12-19 Years: United States, 1999-2006 URL: https://www.cdc.gov/nchs/products/databriefs/db37.htm [accessed 2019-01-19] [WebCite Cache ID 75XobD4H1]
2. McCartt AT, Shabanova V, Leaf WA. Driving experience, crashes and traffic citations of teenage beginning drivers. Accid Anal Prev 2003 May;35(3):311-320. [doi: 10.1016/s0001-4575(02)00006-4] [Medline: 12643948]
3. Curry AE, Pfeiffer MR, Durbin DR, Elliott MR, Kim KH. AAA Foundation for Traffic Safety. 2014. Young Driver Crash Rates in New Jersey by Driving Experience, Age, and License Phase URL: https://aaafoundation.org/young-driver-crash-rates-new-jersey-driving-experience-age-license-phase/ [accessed 2019-03-06] [WebCite Cache ID 76fseBsNP]
4. Chapman EA, Masten SV, Browning KK. Crash and traffic violation rates before and after licensure for novice California drivers subject to different driver licensing requirements. J Safety Res 2014 Sep;50:125-138. [doi: 10.1016/j.jsr.2014.05.005] [Medline: 25142369]
5. Curry AE, Hafetz J, Kallan MJ, Winston FK, Durbin DR. Prevalence of teen driver errors leading to serious motor vehicle crashes. Accid Anal Prev 2011 Jul;43(4):1285-1290. [doi: 10.1016/j.aap.2010.10.019] [Medline: 21545856]
6. McDonald CC, Curry AE, Kandadai V, Sommers MS, Winston FK. Comparison of teen and adult driver crash scenarios in a nationally representative sample of serious crashes. Accid Anal Prev 2014 Nov;72:302-308. [doi: 10.1016/j.aap.2014.07.016] [Medline: 25103321]
7. McKnight AJ, McKnight AS. Young novice drivers: careless or clueless? Accid Anal Prev 2003 Nov;35(6):921-925. [doi: 10.1016/s0001-4575(02)00100-8] [Medline: 12971927]

8. Nef T, Müri RM, Bieri R, Jäger M, Bethencourt N, Tarnanas I, et al. Can a novel web-based computer test predict poor simulated driving performance? a pilot study with healthy and cognitive-impaired participants. J Med Internet Res 2013 Oct 21;15(10):e232 [FREE Full text] [doi: 10.2196/jmir.2943] [Medline: 24144946]
9. Walshe EA, Oppenheimer N, Kandadai V, Winston FK. 2011-2018 Annual Meeting Compendium of Papers Archive. 2019. Comparison of Virtual Driving Test Performance and On-Road Examination for Licensure Performance: A Replication Study URL: http://amonline.trb.org/68387-trb-1.4353651/t0016-1.4367835/1346-1.4368021/19-01210-1.4362689/19-01210-1.4368040?qr=1 [accessed 2019-03-12] [WebCite Cache ID 76oseTRiu]
10. Diagnostic Driving. 2018. How Does Ready-Assess Work? URL: http://www.diagnosticdriving.com/how-readyassess-works [accessed 2019-03-12] [WebCite Cache ID 76osu6cIe]
11. The Orange County Independent. 2017. Drivers License Test Ends in Spectacular Crash at Costa Mesa DMV URL: http://theorangecountyindependent.com/2017/01/drivers-license-test-ends-in-horrendous-crash-at-costa-mesa-dmv/ [accessed 2019-03-12] [WebCite Cache ID 76owL8Fjj]
12. Osborne M. ABC News. 2018. Teen Taking Driver's Exam Crashes Into Test Office URL: https://abcnews.go.com/US/teen-taking-drivers-exam-crashes-test-office/story?id=53998272 [accessed 2019-03-12] [WebCite Cache ID 76owvonTJ]
13. Hooper B. United Press International. 2018. Student Driver Crashes Into DMV During Road Test URL: https://www.upi.com/Student-driver-crashes-into-DMV-during-road-test/6151522848041/ [accessed 2019-03-12] [WebCite Cache ID 76oyV0rb6]
14. McDonald CC, Kandadai V, Loeb H, Seacrist TS, Lee YC, Winston X, et al. Simulated Driving Assessment (SDA) for teen drivers: results from a validation study. Inj Prev 2015 Jun;21(3):145-152 [FREE Full text] [doi: 10.1136/injuryprev-2014-041480] [Medline: 25740939]
15. McDonald CC, Tanenbaum JB, Lee Y, Fisher DL, Mayhew DR, Winston FK. Using crash data to develop simulator scenarios for assessing novice driver performance. Transp Res Rec 2012;2321(1):73-78 [FREE Full text] [doi: 10.3141/2321-10] [Medline: 23543947]
16. CrashStats. 2019. National Motor Vehicle Crash Causation Survey: Report to Congress URL: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811059 [accessed 2019-03-06] [WebCite Cache ID 76fyy4afX]
17. Diagnostic Driving. Ready-Assess for Driver's Licensing Centers URL: http://www.diagnosticdriving.com/for-drivers-license-programs [accessed 2019-03-06] [WebCite Cache ID 76g049FV1]
18. Verster JC, Roth T. Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). Int J Gen Med 2011;4:359-371 [FREE Full text] [doi: 10.2147/IJGM.S19639] [Medline: 21625472]
19. ten Holt GA, Reinders MJ, Hendriks EA. Intelligent Systems. 2007. Multi-Dimensional Dynamic Time Warping for Gesture Recognition URL: http://msp.ewi.tudelft.nl/sites/default/files/DTW-vASCI.pdf [accessed 2019-02-20] [WebCite Cache ID 76KrBCTCD]
20. Rangsit PS. Semantic Scholar. Generalized Dynamic Time Warping for Multiple Multidimensional Sequence Alignment URL: https://pdfs.semanticscholar.org/1b7f/0f25361d8dfcfea6d4e72a64f80b662272ef.pdf [accessed 2019-02-20] [WebCite Cache ID 76LBMurZ5]
21. Schubert E, Rousseeuw PJ. ArXiv. 2018. Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms URL: https://arxiv.org/abs/1810.05691 [accessed 2019-02-20]
22. Weka-SourceForge. Class SimpleLogistic URL: http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/SimpleLogistic.html [accessed 2019-02-20] [WebCite Cache ID 76Krd0PWw]
23. Weka-SourceForge. Class LibSVM URL: http://weka.sourceforge.net/doc.stable/weka/classifiers/functions/LibSVM.html [accessed 2019-02-20] [WebCite Cache ID 76KrjjuxG]
24. Fan RE, Chen PH, Lin CJ. Working set selection using second order information for training support vector machines. J Mach Learn Res 2005;6:- [FREE Full text] [doi: 10.5555/1046920.1194907]
25. Hsu CW, Chang CC, Lin CJ. National Taiwan University: Department of Computer Science. 2016. A Practical Guide to Support Vector Classification URL: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [accessed 2019-02-21] [WebCite Cache ID 76L6iXviD]
26. Tape TG. The Darwin Web Server. The Area Under an ROC Curve URL: http://gim.unmc.edu/dxtests/ROC3.htm [accessed 2019-03-12] [WebCite Cache ID 76ozHQTiq]
27. Eamonn; E, Lin J, Truppel W. Clustering of Time Series Subsequences is Meaningless: Implications for Previous and Future Research. In: Proceedings of the Third IEEE International Conference on Data Mining. 2003 Presented at: ICDM'03; November 22, 2003; Melbourne, FL, USA, USA URL: http://www.cs.ucr.edu/~eamonn/meaningless.pdf [doi: 10.1109/icdm.2003.1250910]
28. Stein G. University of Central Florida. 2009. Falconet: Force-Feedback Approach For Learning From Coaching And Observation Using Natural And Experiential Training URL: https://stars.library.ucf.edu/etd/3988/ [accessed 2019-02-21] [WebCite Cache ID 76MTo0g7E]
29. Achtert E, Böhm C, Hans-Peter K, Kröger P. Robust, Complete, and Efficient Correlation Clustering. In: Proceedings of the Seventh SIAM International Conference on Data Mining. 2007 Presented at: SIAM'07; April 26-28, 2007; Minneapolis,

Minnesota, USA URL: https://pdfs.semanticscholar.org/c45c/baf6f4c7f066a556a6370c144d030311ab3b.pdf [doi: 10.1137/1.9781611972771.37]
30. Cocos A, Callison-Burch C. Clustering Paraphrases by Word Sense. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016 Presented at: NAACL HLT'16; June 12-17, 2016; San Diego, California p. 1463-1472 URL: http://www.aclweb.org/anthology/N16-1172 [doi: 10.18653/v1/n16-1172]
31. Bingham C, Ehsani JP. The relative odds of involvement in seven crash configurations by driver age and sex. J Adolesc Health 2012 Nov;51(5):484-490. [doi: 10.1016/j.jadohealth.2012.02.012] [Medline: 23084170]
32. Dembélé D, Kastner P. Fuzzy C-means method for clustering microarray data. Bioinformatics 2003 May 22;19(8):973-980. [doi: 10.1093/bioinformatics/btg119] [Medline: 12761060]
33. Harb H, Makhoul A, Couturier R. An enhanced K-means and ANOVA-based clustering approach for similarity aggregation in underwater wireless sensor networks. IEEE Sensors J 2015 Oct;15(10):5483-5493 [FREE Full text] [doi: 10.1109/jsen.2015.2443380]

Abbreviations

AUC: area under the curve
CHOP: Children’s Hospital of Philadelphia
FPR: false positive rate
ML: machine learning
MVC: motor vehicle crash
ORE: on-road examination
Penn: University of Pennsylvania
RBF: radial basis function
ROC: receiver operator characteristic
RR: relative risk
SDA: simulated driving assessment
SME: subject matter expertise
SVM: support vector machine
TPR: true positive rate
VDT: virtual driving test

Edited by G Eysenbach; submitted 12.03.19; peer-reviewed by K Goniewicz, N Levitz; comments to author 17.09.19; revised version received 11.11.19; accepted 16.12.19; published 18.06.20.

Please cite as:
Grethlein D, Winston FK, Walshe E, Tanner S, Kandadai V, Ontañón S
Simulator Pre-Screening of Underprepared Drivers Prior to Licensing On-Road Examination: Clustering of Virtual Driving Test Time Series Data
J Med Internet Res 2020;22(6):e13995
URL: https://www.jmir.org/2020/6/e13995
doi:10.2196/13995
PMID:32554384

Original Paper

Using Virtual Reality to Improve Apathy in Residential Aged Care: Mixed Methods Study

Dimitrios Saredakis[1], BPsych; Hannah AD Keage[1], PhD; Megan Corlis[2], RN, BAppSc; Tobias Loetscher[1], PhD

[1]University of South Australia, Adelaide, Australia

[2]Helping Hand Aged Care, Adelaide, Australia

Corresponding Author:
Dimitrios Saredakis, BPsych
University of South Australia
St Bernards Road, Magill
South Australia
Adelaide, 5072
Australia
Phone: 61 08 8302 4083
Email: dimitrios.saredakis@mymail.unisa.edu.au

Abstract

Background: Apathy is a common symptom in neurological disorders, including dementia, and is associated with a faster rate of cognitive decline, reduced quality of life, and high caregiver burden. There is a lack of effective pharmacological treatments for apathy, and nonpharmacological interventions are a preferred first-line approach to treatment. Virtual reality (VR) using head-mounted displays (HMDs) is being successfully used in exposure- and distraction-based therapies; however, there is limited research on using HMDs for symptoms of neurological disorders.

Objective: This feasibility study aimed to assess whether VR using HMDs could be used to deliver tailored reminiscence therapy and examine the willingness to participate, response rates to measures, time taken to create tailored content, and technical problems. In addition, this study aimed to explore the immediate effects between verbal fluency and apathy after exposure to VR.

Methods: A mixed methods study was conducted in a sample of older adults residing in aged care, and 17 participants were recruited. Apathy was measured using the Apathy Evaluation Scale (AES), and verbal fluency was used as a proxy measure of improvements in apathy and debriefing interviews to assess feedback from participants. Side effects that can occur from using HMDs were also measured.

Results: We recruited participants from a high socioeconomic status setting with a high education level, and the participation rate was 85% (17/20); most responses to measures were positive. Access to a wide range of freely available content and the absence of technical difficulties made the delivery of a VR reminiscence intervention highly feasible. Participants had improved semantic scores (t_{14}=−3.27; *P*=.006) but not phonemic fluency scores (t_{14}=0.55; *P*=.59) immediately after the intervention. Those with higher levels of apathy demonstrated the greatest cognitive improvements after a VR reminiscence experience, which was indicated by a strong positive relationship between the AES and semantic verbal fluency change scores postminus pre-VR (*r*=0.719; 95% CI 0.327 to 0.900; *P*=.003). All participants enjoyed the experience despite 35% (6/17) of participants experiencing temporary side effects.

Conclusions: This study provides initial evidence that it is feasible to use VR with HMDs for therapy to treat symptoms of apathy in older adults in residential aged care. However, there is a need to closely monitor the side effects of HMD use in older adults. Further research is needed using an active control group to compare the use of VR with traditional forms of reminiscence therapy.

(J Med Internet Res 2020;22(6):e17632) doi:10.2196/17632

KEYWORDS

reminiscence; head-mounted display; apathy; cognitive aging; dementia; residential facilities; virtual reality

Introduction

Background

Apathy is one of the most common symptoms of neurological and psychiatric disorders such as dementia [1-3]. A lack of motivation is the defining feature of apathy, resulting in a reduction of goal-directed behavior [4]. Apathy contributes to a poorer quality of life and a faster rate of cognitive decline [5,6]. In addition, the risk of mortality in those with apathy increases 3-fold compared with those without apathy [7]. Those caring for people with apathy are also affected by reported higher caregiver burden compared with other neuropsychiatric symptoms [8,9].

A high prevalence of apathy of up to 84% has been reported in older adults in residential aged care with cognitive decline [10]. Healthy older adults can also experience symptoms of apathy in varying degrees [11,12]. Significant changes in life circumstances, for example, surviving natural disasters or being institutionalized, may reduce motivation, and increase symptoms of apathy [13]. Changes in specific brain regions have been associated with apathy, including frontal-subcortical structures [14,15]. However, the mechanisms of apathy are not fully understood.

Although symptoms of apathy may overlap with those of depression [16], apathy can also present as a separate clinical symptom [17], and there is increasing evidence supporting this distinction [18,19]. Symptoms of apathy can also differ depending on the type of dementia, with more severe symptoms found in the behavioral variant of frontotemporal dementia [20]. With little evidence supporting pharmacological treatment for apathy [21] and the side effects of current medications [22], alternative interventions are required.

Nonpharmacological approaches have demonstrated evidence in improving the levels of apathy, including music, art, cognitive stimulation, and reminiscence therapy [23]. Reminiscence therapy is commonly used in residential aged care and involves the discussion of a person’s past experiences involving the use of items including photographs, magazines, or music [24]. It can help maintain a person’s sense of identity and improve feelings of self-worth [25,26], an important aspect of improving apathy.

A recent Cochrane review found reminiscence therapy to be effective in some cases [27]. However, the effect sizes found were small and inconsistent. Little or no differences in cognitive outcomes were commonly found when comparing reminiscence therapy–treated groups with nontreatment groups (eg, standardized mean difference 0.11, 95% CI 0.00 to 0.23) [27]. Improving the delivery of tailored reminiscence therapy by increasing how realistic it is may increase the effectiveness of the therapy and the likelihood of attending sessions [28].

The use of technology in reminiscence therapy can provide access to a wide range of content via the internet and make it easy to tailor an intervention specifically for a person [29,30]. In a within-group study comparing the creation of a digital life storybook with a conventional life storybook for people with dementia, additional improvements were seen in the quality of life and autobiographical memory performance when the digital life storybook was used [30]. The traditional version in this study consisted of tangible prompts, including photos and other memorabilia, and the digital version consisted of a movie created and burned onto a DVD for playing on television. iPads have also been used in the delivery of individualized reminiscence therapy with positive results, including improvements in well-being [31,32]. Significantly higher levels of initiating speech and engagement were found using a touchscreen monitor compared with traditional reminiscence therapy using artifacts, including newspapers, books, and photos [33]. Virtual reality (VR) in the form of large-screen displays has also been used and provides an immersive experience [34]. The level of immersion can vary among technologies. iPads or computer screens are classified as nonimmersive, whereas large-screen displays are considered semi-immersive [35]. An increase in the level of immersion can increase the level of presence (the degree to which a person feels they are in the virtual environment) [36]. During a navigation task, an increased sense of presence using highly immersive 3D screens while wearing stereo glasses has been found to recruit additional brain areas compared with less immersive 2D screens [37].

VR using head-mounted displays (HMDs) provides a fully immersive, highly engaging, and realistic experience with the user closed off from external distractions. The use of HMDs is not limited to gaming and has infiltrated many areas, including medical training, rehabilitation, and therapy [38-41]. Advances in technology, production, and reduced costs have made HMDs widely accessible, with little training required to implement and use them. In patients who had previously not reported clinical benefits from other forms of psychotherapy, the use of VR using HMDs has been successful in treating post-traumatic stress disorder (PTSD), with a significant decrease in PTSD symptoms being found while comparing pretreatment symptoms with posttreatment symptoms (d=1.17) [42]. Other successful interventions using HMDs include reducing anxiety [43] and pain management of burn victims during wound care [44,45].

Recent research has found that older adults, including those with cognitive decline, can tolerate using HMDs and has reported positive experiences [46-49]. There has also been success in using HMDs for memory training in older adults with significant improvements in memory tests (d=.7) compared with a control group receiving music therapy [50]. Other uses of HMDs, including in older adults, have shown an increase in the performance of daily living activities in stroke victims [51] and success with vestibular rehabilitation in older adults with mild cognitive impairment [52]. These findings of HMD use in older adults suggest that expansion into other areas is possible. Research into the use of VR using HMDs in therapy for symptoms of neurological disorders is limited.

In addition to the lack of motivation in apathy [53], executive dysfunction has also been found to be associated with apathy [54,55]. Impairment in executive function is a key cognitive mechanism of apathy. Apathy is negatively associated with the performance of verbal fluency tasks [56]. Similarly, those with dementia and apathy have shown reduced performance on verbal fluency tasks, compared with those without apathy [57]. Specifically, associations have been reported between apathy and semantic fluency but not phonemic fluency [56,58]. Performance in verbal fluency tasks indicates executive control [59] and initiation [60]. These executive functions of control and initiation are reduced in those with apathy [61,62]. Apathetic participants demonstrated a preference for a VR task over a pen and paper task, as it provided a more engaging experience [63]. In addition, VR (using a projector screen with stereo glasses) has demonstrated the ability to stimulate autobiographical memories [34]. Therefore, VR may improve the performance of executive functions, both directly and indirectly.

Although there is evidence supporting the acceptability of using HMDs in older adults as previously mentioned, some people do experience side effects from using HMDs, including the occurrence of motion sickness such as that induced by air, sea, or land travel. Research on sickness symptoms while using HMDs in older adults has reported conflicting outcomes [64-66]. Many companies are starting to offer VR services for older adults in residential aged care [67-69]. However, the occurrence and prevalence of side effects of using VR HMDs with this age group are uncertain [70]. With the increased use of VR using HMDs in many areas, including older adults, awareness of any side effects from using this technology has become essential.

Objectives

This study aimed to evaluate the feasibility of using tailored content created for viewing in a VR HMD to deliver reminiscence therapy to older adults in residential aged care. Along with establishing the feasibility of this study, the following parameters were also examined: availability and willingness to participate, response rates to measures used, the time needed to create tailored content, and technical problems. Verbal fluency was used as a measure of immediate effects on executive function after exposure to tailored content in VR, and associations with verbal fluency and apathy were also examined. A debriefing interview was used to assess enjoyment and obtain feedback from participants. Side effects of using HMDs were also examined using the Simulator Sickness Questionnaire (SSQ) combined with feedback from the debriefing interview.

Methods

Participants

Participants were recruited from a residential aged care facility in Adelaide, Australia. Potentially suitable participants were identified by senior staff at the residential aged care facility following the inclusion and exclusion criteria. A list of identified participants was then given to the researcher, and the researcher approached participants to request their participation in the study. The current health status of each participant was obtained from the nursing staff on duty before the participants were contacted. Out of 20 participants who were requested to participate in this study, 3 declined; 10 women and 7 men, between the ages of 72 and 95 years with a mean age of 87.3 years (SD 6.3), agreed to participate. All the participants were Australian, of whom 11 were born in Australia, 2 in the United Kingdom, 2 in Africa, and 2 in Europe. The participants born in the United Kingdom, Africa, and Europe had spent most of their lives in Australia. The participants came from a high socioeconomic background and most were well educated; the mean years of education was 13 (SD 3). According to the Psychogeriatric Assessment Scale (PAS), 10 participants had no or minimal cognitive impairment, 3 presented with mild impairments, and 4 had moderate impairments [71]. The mean score for the Apathy Evaluation Scale (AES) was 31.5 (SD 6.3). Using a cutoff score of 36.5 [72], 3 participants met the criteria for a diagnosis of apathy with the highest score being 46; 5 participants scored between 34 and 36. Medication history was available for 14 participants. The most common medication used was for depressive symptoms in 6 out of 14 participants (43%). Participants were excluded if their score on the PAS was 16 or higher indicating severe impairment; had significant neurological disorders; had other conditions, including agitation and aggression at a level that would make assessment difficult; or if their vision could not be corrected by glasses or glass frames that could not fit into the HMD. Informed consent was obtained from all the participants, and no honorarium was offered for participation. Ethics approval was obtained from the University of South Australian Human Research Ethics Committee in accordance with the Australian National Statement on Ethical Conduct in Human Research (2007).

Materials

Virtual Reality Apparatus

The Oculus Go HMD was used [73] to deliver the VR experience to participants. This is a standalone HMD that does not require tethering to a computer or the use of a mobile phone.

Virtual Reality Content

YouTube VR (Google LLC) was used for the playback of 360-degree videos. This is an app that is specifically used for viewing YouTube videos in VR HMDs. All 360-degree videos were downloaded to the HMD to reduce reliance on the internet and improve the reliability of playback. Wander (Parkline Interactive) was used to navigate to places of interest tailored to the participant. This app uses data from Google Street View and has the feature of saving places visited for future access.

Questionnaire for Establishing Virtual Reality Content

An interview was conducted to develop tailored content for each participant and was based on reminiscence therapy guidelines [24]. Questions included the following: “What significant memories do you have from your childhood?” “What significant life events do you have positive memories of?” and “What memories do you have from working, including details about your first job, and work in the home?”

Psychogeriatric Assessment Scale

Cognitive status was obtained from the current records of the residential aged care facility using the PAS [74]. The PAS includes both a subject and informant interview and provides a summary of functioning related to cerebrovascular disease, cognitive status, and behavioral disturbance that can occur in dementia [74]. Scores range from 0 to 21 with a score of ≥4 indicating likely cognitive impairment [75] and the risk of further cognitive decline [76].

Apathy Evaluation Scale Self-Rated Version

Apathy was measured using the AES [4]. This was included to measure the participants' current level of apathy. The AES contains 18 items measured on a scale ranging from *not at all*, *slightly*, *somewhat*, and *a lot*. The AES scores range from 18 to 72, with higher scores indicating a higher level of apathy. Rating is based on the previous 4 weeks. The self-rated version of the AES has been found to have satisfactory test-retest reliability of .76 and good internal consistency (α=.86) [4]. The internal consistency for this study was α=.77. Questions included, "Are you interested in having new experiences?" and "Do you spend time doing things that interest you?"

Simulator Sickness Questionnaire

Side effects were measured using the SSQ [77]. This questionnaire is widely used in VR-related research [78]. It consists of 16 items that assess the symptoms in the VR environment using a 4-point rating scale from 0 (indicating *no symptoms*) to 3 (indicating *severe symptoms*). The SSQ has 3 subscales, including nausea, oculomotor, and disorientation, and has demonstrated good internal consistency (α=.87) [79]. Internal consistency for this study was α=.69. Both the total and subscale SSQ scores require weighting. The subscale scores were weighted as follows: nausea 9.54, oculomotor 7.58, and disorientation 13.92. Total scores were calculated by adding the 3 unweighted subscale scores and multiplying by 3.74. The 3 subscales include overlapping symptoms, for example, the symptom of *blurred vision* overlaps both the oculomotor and disorientation subscales. Symptoms measured included *headache*, *nausea*, *sweating*, and *eye strain*.

Slater-Usoh-Steed Presence Questionnaire

The Slater-Usoh-Steed Presence Questionnaire (SUS) was used to assess how realistic participants thought the VR environment was [80]. The SUS consists of 6 items scored from 1 to 7. Questions include "To what extent were there times during the experience when the virtual environment was reality for you?" and "During the time of your experience, did you often think to yourself that you were actually in the virtual environment?" This scale has demonstrated acceptable internal consistency (α=.75) [81]. Internal consistency for this study was α=.74.

Verbal Fluency

Both phonemic and semantic verbal fluency tasks were administered to the participants [82]. In the phonemic verbal fluency task, participants had to name as many words as possible starting with the letter *F* or *P*. In the semantic verbal fluency task, participants had to list as many words as possible in the category of either *animals* or *fruit/vegetables*. The tasks were timed for 1 min each.

Expectations/Enjoyment Measure

Participants were asked about their expectations before the VR experience "What are your expectations about the VR experience, in reference to overall enjoyment?" After the VR experience, they were asked, "How much did you enjoy the experience?" Each question was given a rating ranging from 1 to 10.

Debriefing Questionnaire

A debriefing interview consisting of 8 questions at the end of the VR session (session 2) assessed the feedback from the participants. Interview questions included "Did you find the experience enjoyable?" and "What did you like about the experience?"

Procedure

The study consisted of 2 sessions, with each session lasting for approximately 60 min. Before consent, to ensure that the participants were able to see images satisfactorily in the headset, a short demonstration was performed where participants viewed images in the HMD using the Wander app. Once it was established that the participants could satisfactorily see the images in the headset, and they consented to participate in the study, an appointment was made for the first session. In the first session, basic demographic data were collected. Participants were then asked a series of questions (questionnaire for establishing VR content) to establish topics for sourcing content to be viewed in the HMD. Finally, the participants completed the AES.

Once the researcher had sourced content based on data gathered from the interview in the first session for viewing in the HMD (approximately 3 days apart), an appointment was made for the second session. Before and after the VR session, 2 verbal fluency tests (phonemic and semantic), expectations/enjoyment rating, and the SSQ were administered. Alternate forms of verbal fluency tests were counterbalanced for each participant. Although the SSQ was designed to be used postexposure to VR, it was decided to follow the precondition of screening participants who may have health-related conditions [77]. The sourcing of content was completed between the first and second sessions and took approximately 90 min for each participant. Participants viewed the VR content that was timed for 20 min. The researcher determined the order of the content being viewed, and during the VR session, a conversation occurred between the participant and the researcher about the participant's past experiences with regard to the content. Participants were not given access to the controllers, and all navigation was performed by the researcher. This was done to minimize frustration and provide a seamless experience for each participant. The HMD image was mirrored to a laptop to enable the researcher to see what the participant was viewing and to allow the researcher to navigate within the apps. Participants remained seated during the VR content exposure, as standing has been found to increase the risk of motion sickness when using HMDs [83]. Remaining seated also reduced the risk of falls while wearing an HMD. A swivel chair was used to enable participants to turn freely when viewing the VR content, thereby limiting neck movement. One participant was restricted to a mobile chair, and the VR content was delivered while they were lying down at a 45-degree angle in the chair. The researcher turned the mobile chair during VR exposure to enable the participant to view the 360-degree videos. Another participant with mobility issues was seated in a lounge-type chair in their room. After VR exposure, the SUS was administered as a measure of presence, followed by the debriefing questionnaire.

Results

Separate Bonferroni corrections for multiple comparisons were applied for each family of tests. For the 2 fluency tasks, alpha was set at <.025 (.05/2). For the 4 SSQ scores, alpha was set at <.013 (.05/4). For the pre-expectations and enjoyment scale, alpha was set at <.05 (.05/1). Assumptions of normality using the Shapiro-Wilk test found both oculomotor and disorientation subscale SSQ scores to significantly deviate from a normal distribution; therefore, the Wilcoxon signed-rank test was performed for all SSQ comparisons. Two-tailed paired samples *t* tests were used for the remaining comparisons. Means and standard deviations for all pre- and postmeasures are shown in Table 1. Two-tailed paired sample *t* tests revealed that semantic verbal fluency scores were significantly higher post-VR, which was the only significant difference (Table 1).

Neither pre- nor post-VR comparison between phonemic and semantic fluency reached statistical significance in a two-tailed paired samples *t* test (pre-VR: mean 0.40 (SD 5.10); 95% CI −2.42 to 3.22; t_{14}=0.30; *P*=.77. post-VR: mean −2.47 (SD 5.21); 95% CI −5.35 to 0.42; t_{14}=−1.83; *P*=.09).

Bivariate (Pearson) correlations between the AES scores with either phonemic or semantic verbal fluency taken at baseline were not significant (AES and phonemic: *r*=0.309; 95% CI −0.241 to 0.709; *P*=.26; AES and semantic: *r*=−0.104; 95% CI −0.585 to 0.432; *P*=.71).

Bivariate (Pearson) correlations between the AES scores taken at baseline and phonemic verbal fluency change scores postminus pre-VR were not significant (*r*=−0.195; 95% CI −0.643 to 0.352; *P*=.49. There was a strong positive relationship between the AES and semantic verbal fluency change scores postminus pre-VR (*r*=0.719; 95% CI 0.327 to 0.900; *P*=.003; Figure 1).

Correlations between the AES scores and either phonemic or semantic change scores were significant (*P*=.03). Statistical differences between correlations were calculated using the Steiger method [84]. Figure 2 shows a breakdown of the SSQ scores for each participant pre- and post-VR using SSQ score cutoffs [85]. Figure 3 shows a summary of the total and subscale SSQ scores. Figure 4 shows a breakdown of the pre-expectation/postenjoyment scores for each participant.

Table 1. Means and standard deviations for all pre- and post-VR measures with statistics and effect size.

Measures	Pre-VR[a], mean (SD)	Post-VR, mean (SD)	*t* statistic (*df*)	*P* value	*d* value
Phonemic verbal fluency	10.13 (3.78)	9.73 (3.81)	0.55 (14)	.59	0.142
Semantic verbal fluency	9.73 (4.10)	12.20 (4.54)	−3.27 (14)	*.006* [d]	−0.843
SSQ total[b]	12.22 (13.07)	13.46 (11.38)	24.00[c]	.45	−0.230
SSQ nausea	15.90 (15.99)	17.17 (17.74)	14.00[c]	.62	−0.107
SSQ oculomotor	18.70 (21.98)	20.72 (20.75)	8.50[c]	.40	−0.174
SSQ disorientation	7.42 (15.67)	12.06 (14.76)	1.00[c]	.10	−0.408
Expectations/enjoyment	6.80 (2.18)	8.07 (1.87)	−2.02 (14)	.06	−0.520

[a] VR: virtual reality.

[b] SSQ: Simulator Sickness Questionnaire.

[c] Denotes the Wilcoxon signed-rank test. The remaining comparisons are paired samples two-tailed *t* tests.

[d] Significant value shown in italics.

Figure 1. Correlation between scores on the Apathy Evaluation Scale at baseline and postminus pre–virtual reality session semantic and phonemic verbal fluency. Two participants scored the same in the semantic verbal fluency correlation.

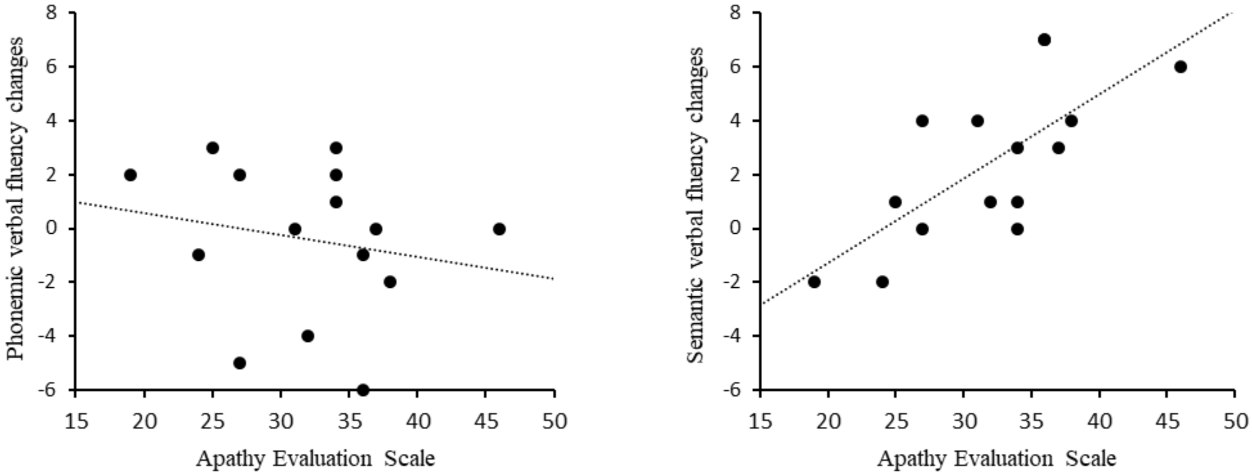


Figure 2. Pre– and post–Simulator Sickness Questionnaire scores with each line representing 1 participant. Shaded areas indicate Simulator Sickness Questionnaire score cutoffs of <5, 5 to 10, 10 to 15, 15 to 20, and >20. VR: virtual reality.

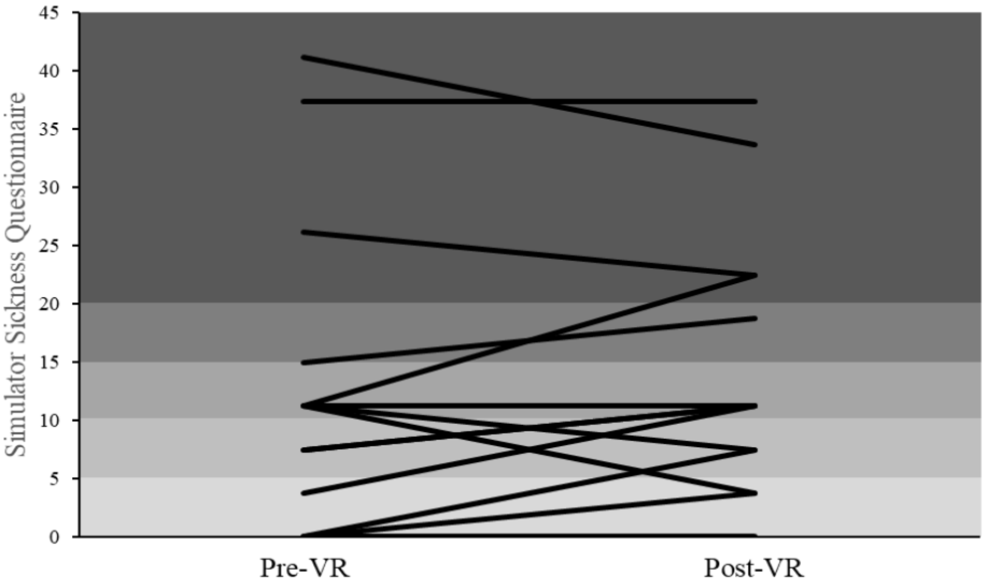
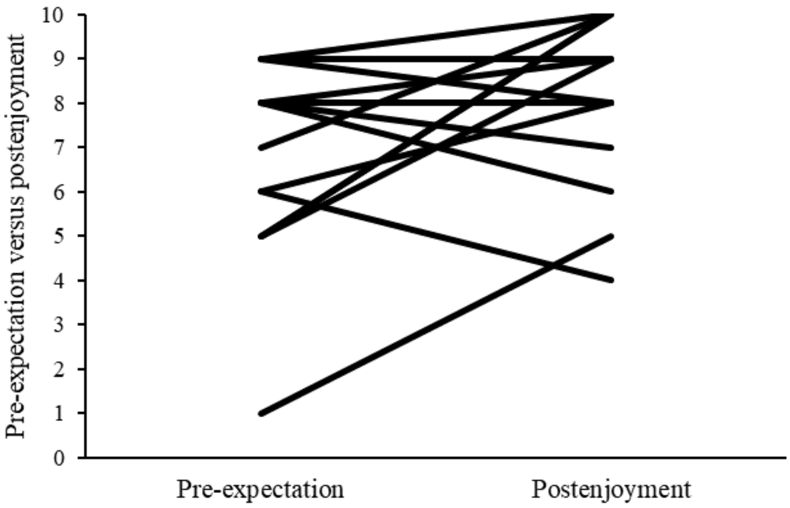


Figure 3. Distribution of Simulator Sickness Questionnaire total and subscale scores at pre- and post–virtual reality intervention, indicated by boxplots within violin plots, white dots represent medians.

Figure 4. Pre–virtual reality expectations versus post–virtual reality enjoyment scores with each line representing one participant.



Most participants had some difficulty with the SUS questionnaire; question number 5 had to be repeated or further explained. Four participants did not attempt answering question number 5, and 2 participants could not complete any of the questions. Therefore, data from the SUS were not analyzed.

After the VR session, the participants were asked a set of 8 questions to gain feedback about their experiences. Although some of the questions could be answered with a yes or no response, participants were given the opportunity to elaborate on their answers, if required.

All the study participants found the experience enjoyable. When asked what they liked about the experience, 11 participants said they enjoyed reminiscing, 4 participants reported it was nice to do something different, 1 participant reported how realistic the experience was, and finally, 1 participant liked the colors and clarity of the images shown. Comments included “liked the accuracy, feeling of being there and sites that I wanted to see again,” “taken back to somewhere that I know I will never see again,” “better than looking at a computer,” and “I had a feeling of being there.”

In reference to acceptability, 1 participant reported that the HMD was heavy; however, this was rectified by adjusting the head strap. All remaining participants found everything acceptable. When asked if they would do it again, all responded yes, with 1 participant stating, “only for research purposes.”

There were instances where participants reported negative symptoms or side effects. This included 1 participant with discomfort around the cheekbone and forehead areas. Another participant reported dizziness and the feeling that they were “going to fall off the edge;” a further 2 participants reported the feeling that they were going to fall off the edge. One participant reported eyestrain, and another participant reported nausea 10 min after the VR session, which subsided within 10 min. Therefore, 6 out of 17 participants reported some type of negative symptoms or side effects.

Participants were asked to offer suggestions to improve their experience. One participant stated that they would have liked to use the controller to navigate in the virtual environment on their own; another participant suggested having some type of interaction; a third participant said it would have been good to “show the difference between then and now,” referring to the Wander app; and finally, one participant would have preferred to see the places that they had not been to before. All participants said they would recommend the VR experience to a friend.

Finally, the participants were asked what they thought about the questionnaires. A total of 12 participants reported that they were all acceptable; 1 participant did not like them without giving a reason; 1 participant could not remember enough to comment; 1 participant said there were too many questions; 1 participant thought there was not a lot of relevance; and 1 participant said that the pre-VR measures did “cause some tension.”

Discussion

This study found that VR using HMDs to deliver reminiscence therapy is feasible and may be beneficial for aged care residents with apathy. Of the 20 participants who were requested to participate, 17 agreed, indicating a willingness of residents in aged care to participate in research using VR technology. Therefore, the number of available participants for this study was easily achieved. The 3 participants who declined to participate were not interested. It is not known how often exclusion criteria had to be applied, as participants were selected by senior staff at the residential aged care facility following the exclusion criteria. A challenge for future research in apathy is a possible reluctance of those with high levels of apathy to participate because of a lack of interest and motivation, similar to reduced involvement in therapeutic activities found in those with higher levels of apathy [86]. The response rates to measures used in this study were positive. The SUS was the only measure that caused difficulties in understanding the questions, with 2 participants not completing any of the questions. The SUS was chosen because of the small number of items (6) and quick completion time compared with other measures of presence [81] to reduce participant burden. It was validated in a younger sample of university-based students with a technical background, who had no difficulties in completing the questionnaire [81].

As per the qualitative feedback received by this study, older adults may have difficulties in understanding some questions of the SUS. All other measures were fully completed, and all participants adhered to the procedure of the research. It should be noted that the reliability values for measures reported in the literature may not generalize to this study because of sample differences, that is, older adults with neurological and other chronic diseases.

The time taken to create content varied depending on the participant's previous experiences. In this study, the researcher prepared the content before the VR session. It was found that by using readily available apps, the content could easily be sourced for all participants for a VR exposure time of 20 min. This finding is of importance for this study, as individualized reminiscence therapy can be resource intensive [87]. The accessibility of content utilizing the apps in this study reduces the resources required for creating content and places fewer demands on staff if applied in aged care lifestyle activities.

In this study, the preparation of VR content took approximately 90 min for each participant. This time can be substantially reduced once a library of 360-degree travel or music videos has been created. In addition, in this study, the content was compiled before the VR session to maximize the amount of time the participant would reminisce. Sourcing of VR content could also be done during a reminiscence session, working together with the participant while they are viewing images in the HMD, and an area for future research. Working together with the participant not only reduces the preparation time but also allows sessions to be tailored and flexible depending on participants' responses [88]. Traditional methods, including making up a life history book, can require substantial resources and are also limited [88]. Using the Wander and YouTube apps reduces the preparation time and provides a wide range of content that can be easily changed. The use of a group reminiscence therapy format that is more generic can be a preferred approach because of the ability to improve social interaction [89] and can also be less resource intensive [90]. However, each person has a unique history, and using individualized content has been demonstrated to be more effective than using generic content [91]. In addition, an individualized approach is person-centered, focusing on a person's strengths, thereby helping maintain a sense of self and identity [92].

No technical problems with conducting the sessions were encountered. To enable successful delivery of the content, it is recommended that the VR headset be mirrored to an external screen and controlled by the person delivering the content. Thus, the focus is on viewing the content rather than trying to learn to navigate or use the controllers. Although 2 participants would have liked to interact in some form with the virtual environment, this study was designed to provide a seamless experience. As the VR exposure consisted of one session, the intention was to reduce any frustration that could occur because of pressing the wrong buttons on the controller. We cannot draw any conclusions about the extent to which aged care residents could navigate within the app and use the controllers themselves, as this was not attempted. However, the feasibility of controller usage and navigation in VR in this population is an avenue for future research for studies conducted over a longer period. Providing the ability to interact allows the user to act independently and may improve the sense of presence. Recent research has found that the use of VR in providing interaction may be an alternative way of delivering stimulation to people with dementia who do not participate in other lifestyle activities [93].

Verbal fluency was used as a proxy measure to assess improvements in apathy. Both semantic and phonemic verbal fluency results were similar at baseline, despite phonemic fluency being known to be more difficult [94]. This could be because of the level of education in this group of participants (mean 13 years, SD 3), as higher education has been correlated with increased performance in phonemic fluency [95]. In pre-VR sessions, both phonemic and semantic fluency scores were not associated with the participants' level of apathy. In post-VR sessions, not only did semantic fluency increase but the results were positively correlated with participants' level of apathy at baseline. This suggests that the greatest cognitive improvements were seen in participants with higher levels of apathy. What is unclear is why these differences were seen only in semantic fluency. Similar improvements in verbal fluency have been reported in a study using traditional reminiscence therapy in a sample of dementia patients [96] in a short number of sessions. However, in this study, increases were seen in both phonemic and semantic fluency, and apathy was not measured. The executive function of initiation required for verbal fluency may be low in persons with apathy [97], and the reliving of autobiographical memories in an immersive environment may have stimulated the process of initiation.

The results from the debriefing questionnaire were positive and in agreement with previous research, finding acceptance of using HMDs in older adults [46-48]. This study further expands the knowledge of acceptability by demonstrating that VR using HMDs can be used to deliver reminiscence therapy with participants of varying levels of cognitive impairment, including minimal, mild, and moderate impairment, as assessed by the PAS. All participants reported enjoying the VR experience, despite 6 participants reporting a side effect. Another indicator of positive response was that all participants stated they would do it again (although 1 said "for research purposes only"). The most common response about what the participants liked about the experience was related to reminiscence and the enjoyment of seeing places from their past, indicating that this is something this age group wants to experience in a VR environment. An important aspect of reminiscence is maintaining a person's sense of identity [87]. As loss of identity can be a key feature of Alzheimer disease that can increase apathy [98], participating in tailored reminiscence therapy may help in maintaining a person's sense of self and identity and therefore improve levels of apathy. The expectations questions asked pre-VR compared with the enjoyment level post-VR, although in a positive direction, failed to reach a significant level. There were 2 participants in pain because of pre-existing conditions during the VR session, which may have contributed to this result.

Although quantitative findings of short-term side effects from VR were not significant, the debriefing questionnaire raised some concerns. Six participants reported side effects, 4 of which were symptoms related to items measured by the SSQ, including nausea, eyestrain, and dizziness. These symptoms dissipated quickly. In one instance where a participant experienced nausea, it occurred 10 min after the VR experience with symptoms subsiding within 10 min. This after-effect highlights the importance of monitoring participants both during and after VR to avoid health and safety implications.

Moreover, 3 participants had the feeling of “falling off the edge” in VR. This symptom was not captured by the SSQ and did create some discomfort, with 2 of the participants requesting to be moved back to their original position, when they had not actually moved. This could not be associated with the type of content. One explanation could be that sometimes the camera shooting the VR content is elevated, and a high sense of presence may have given a realistic feeling of their position in the virtual environment. Another factor that may have contributed to this symptom is the common occurrence of balance problems in older adults with neurological diseases [99]. This symptom will need to be monitored in future studies when working with this population.

The importance of taking pre- and post-SSQ measures was demonstrated in this study, with some participants reporting symptoms before the commencement of VR. This was usually because of pain from existing conditions. It is also interesting to note that in 4 participants, the SSQ scores decreased post-VR; this included a participant with the highest pre-VR SSQ score. This reduction in the SSQ scores could be a similar effect seen in the successful use of VR for distraction-based therapies [44,45]. Despite all participants finding the experience enjoyable, it is important to consider that 6 out of 17 participants (35%) had some type of negative response when viewing the VR content and the importance of monitoring symptoms. When using cutoff scores for the SSQ (negligible symptoms: <5, minimal symptoms: 5-10, significant symptoms: 10-15, concerning symptoms: 15-20, and problematic symptoms: >20 [85]), it can be seen in Figure 2 that the majority of participants scored >5, with 4 participants experiencing problematic symptoms (>20). However, these cutoffs were created with military personnel using flight simulators, and how these cutoff scores relate to the general population and, in the case of this study, older adults is unknown.

This study aimed to assess feasibility; therefore, a control group was not included, and we could not compare results using a traditional reminiscence therapy approach. However, the aim was first to establish if HMDs could be used in this population for reminiscence therapy. Although alternate forms of both phonemic and semantic verbal fluency were used to reduce practice effects, the lack of a control group receiving no VR intervention warrants a caveat when interpreting the results. It must remain open to what degree the pre-post comparisons are the result of the VR exposure versus a pure retest effect. It is also not known how long the verbal fluency improvements would last and if other aspects of apathy were reduced. Anecdotally, staff members providing services to participants involved in the research reported positive mood changes during the research process, and there were also instances where positive changes in behavior were reported by family members of participants involved in the research.

As the self-rated version of the AES was used, it needs to be considered that people with apathy will normally underestimate their condition [4]. Therefore, true scores may be higher than reported. This study did not specifically recruit participants with apathy, and a challenge with future research is the difficulty of recruiting participants who do have apathy because of their lack of interest and motivation.

This study demonstrated that VR using HMDs can be used to deliver tailored content about a person’s past and that residents in aged care enjoyed this experience. The aged care site in this study was in an area of high socioeconomic status, and study participants had a relatively high education level. Higher levels of education have been associated with an increased willingness to participate in research [100]. The sample was also selected by the staff and therefore was not a random sample, possibly introducing a selection bias [101]. How these factors would generalize to a random sample with an average or lower education level is unknown and needs to be considered when determining generalizability as participation rates, adherence to sessions, and therefore feasibility may differ. Immediate effects on cognition for those with higher levels of apathy after 1 session is an interesting finding and highlights the importance of continuing this research. The VR experience can be implemented and incorporated into current lifestyle activities in aged care with the correct protocols in place. VR is being increasingly used in older adults in areas including assessment, cognitive training, and rehabilitation; therefore, it is critical to monitor side effects, particularly after the VR experience, to avoid any health and safety implications and assess whether the benefits of using HMDs outweigh the risk of side effects. This study provides initial evidence that it is feasible to use VR with HMDs for therapy to treat symptoms of apathy in older adults in residential aged care. Further research over multiple sessions using an active control group to compare with traditional reminiscence therapy will help establish the advantages of using VR with HMDs.

Acknowledgments

DS was supported by the Australian Government Research Training Program Scholarship. TL and HK were funded by the National Health and Medical Research Council Boosting and Dementia Research Leadership Fellowships Scheme (GNT1136269 and GNT1135676, respectively).

Conflicts of Interest

None declared.

References

1. Zhao Q, Tan L, Wang H, Jiang T, Tan M, Tan L, et al. The prevalence of neuropsychiatric symptoms in Alzheimer's disease: systematic review and meta-analysis. J Affect Disord 2016 Jan 15;190:264-271. [doi: 10.1016/j.jad.2015.09.069] [Medline: 26540080]
2. Skorvanek M, Rosenberger J, Gdovinova Z, Nagyova I, Saeedian RG, Groothoff JW, et al. Apathy in elderly nondemented patients with Parkinson's disease: clinical determinants and relationship to quality of life. J Geriatr Psychiatry Neurol 2013 Dec;26(4):237-243. [doi: 10.1177/0891988713500587] [Medline: 23970460]
3. Lanctôt KL, Agüera-Ortiz L, Brodaty H, Francis P, Geda Y, Ismail Z, et al. Apathy associated with neurocognitive disorders: recent progress and future directions. Alzheimers Dement 2017 Jan;13(1):84-100. [doi: 10.1016/j.jalz.2016.05.008] [Medline: 27362291]
4. Marin R, Biedrzycki R, Firinciogullari S. Reliability and validity of the apathy evaluation scale. Psychiatry Res 1991 Aug;38(2):143-162. [doi: 10.1016/0165-1781(91)90040-v] [Medline: 1754629]
5. Geda YE, Roberts RO, Mielke MM, Knopman DS, Christianson TJ, Pankratz VS, et al. Baseline neuropsychiatric symptoms and the risk of incident mild cognitive impairment: a population-based study. Am J Psychiatry 2014 May;171(5):572-581 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13060821] [Medline: 24700290]
6. Tierney SM, Woods SP, Weinborn M, Bucks RS. Real-world implications of apathy among older adults: independent associations with activities of daily living and quality of life. J Clin Exp Neuropsychol 2018 Nov;40(9):895-903. [doi: 10.1080/13803395.2018.1444736] [Medline: 29532707]
7. van der Linde RM, Matthews FE, Dening T, Brayne C. Patterns and persistence of behavioural and psychological symptoms in those with cognitive impairment: the importance of apathy. Int J Geriatr Psychiatry 2017 Mar;32(3):306-315. [doi: 10.1002/gps.4464] [Medline: 27017917]
8. Dauphinot V, Delphin-Combe F, Mouchoux C, Dorey A, Bathsavanis A, Makaroff Z, et al. Risk factors of caregiver burden among patients with Alzheimer's disease or related disorders: a cross-sectional study. J Alzheimers Dis 2015;44(3):907-916. [doi: 10.3233/JAD-142337] [Medline: 25374109]
9. Lou Q, Liu S, Huo YR, Liu M, Liu S, Ji Y. Comprehensive analysis of patient and caregiver predictors for caregiver burden, anxiety and depression in Alzheimer's disease. J Clin Nurs 2015 Sep;24(17-18):2668-2678. [doi: 10.1111/jocn.12870] [Medline: 26108739]
10. Zuidema S, Koopmans R, Verhey F. Prevalence and predictors of neuropsychiatric symptoms in cognitively impaired nursing home patients. J Geriatr Psychiatry Neurol 2007 Mar;20(1):41-49. [doi: 10.1177/0891988706292762] [Medline: 17341770]
11. Bonnelle V, Manohar S, Behrens T, Husain M. Individual differences in premotor brain systems underlie behavioral apathy. Cereb Cortex 2016 Feb;26(2):807-819 [FREE Full text] [doi: 10.1093/cercor/bhv247] [Medline: 26564255]
12. Brodaty H, Altendorf A, Withall A, Sachdev P. Do people become more apathetic as they grow older? A longitudinal study in healthy individuals. Int Psychogeriatr 2010 May;22(3):426-436. [doi: 10.1017/S1041610209991335] [Medline: 20003630]
13. Marin RS. Differential diagnosis and classification of apathy. Am J Psychiatry 1990 Jan;147(1):22-30. [doi: 10.1176/ajp.147.1.22] [Medline: 2403472]
14. le Heron C, Holroyd CB, Salamone J, Husain M. Brain mechanisms underlying apathy. J Neurol Neurosurg Psychiatry 2019 Mar;90(3):302-312 [FREE Full text] [doi: 10.1136/jnnp-2018-318265] [Medline: 30366958]
15. Starkstein S, Brockman S. The neuroimaging basis of apathy: empirical findings and conceptual challenges. Neuropsychologia 2018 Sep;118(Pt B):48-53. [doi: 10.1016/j.neuropsychologia.2018.01.042] [Medline: 29410070]
16. Starkstein SE, Ingram L, Garau ML, Mizrahi R. On the overlap between apathy and depression in dementia. J Neurol Neurosurg Psychiatry 2005 Aug;76(8):1070-1074 [FREE Full text] [doi: 10.1136/jnnp.2004.052795] [Medline: 16024880]
17. Levy ML, Cummings JL, Fairbanks LA, Masterman D, Miller BL, Craig AH, et al. Apathy is not depression. J Neuropsychiatry Clin Neurosci 1998;10(3):314-319. [doi: 10.1176/jnp.10.3.314] [Medline: 9706539]
18. Lohner V, Brookes RL, Hollocks MJ, Morris RG, Markus HS. Apathy, but not depression, is associated with executive dysfunction in cerebral small vessel disease. PLoS One 2017;12(5):e0176943 [FREE Full text] [doi: 10.1371/journal.pone.0176943] [Medline: 28493898]
19. Zhu C, Grossman H, Sano M. Why do they just sit? Apathy as a core symptom of Alzheimer disease. Am J Geriatr Psychiatry 2019 Apr;27(4):395-405 [FREE Full text] [doi: 10.1016/j.jagp.2018.12.013] [Medline: 30655032]
20. Kumfor F, Zhen A, Hodges J, Piguet O, Irish M. Apathy in Alzheimer's disease and frontotemporal dementia: distinct clinical profiles and neural correlates. Cortex 2018 Jun;103:350-359. [doi: 10.1016/j.cortex.2018.03.019] [Medline: 29704671]
21. Ruthirakuhan M, Herrmann N, Abraham E, Chan S, Lanctôt KL. Pharmacological interventions for apathy in Alzheimer's disease. Cochrane Database Syst Rev 2018 May 4;5:CD012197 [FREE Full text] [doi: 10.1002/14651858.CD012197.pub2] [Medline: 29727467]
22. Bogdan A, Manera V, Koenig A, David R. Pharmacologic approaches for the management of apathy in neurodegenerative disorders. Front Pharmacol 2019;10:1581 [FREE Full text] [doi: 10.3389/fphar.2019.01581] [Medline: 32038253]

23. Goris ED, Ansel KN, Schutte DL. Quantitative systematic review of the effects of non-pharmacological interventions on reducing apathy in persons with dementia. J Adv Nurs 2016 Nov;72(11):2612-2628. [doi: 10.1111/jan.13026] [Medline: 27221007]
24. Westphal A, Callega D, LoGiudice D, Lautenschlager N. University of Melbourne. 2017. Using Reminiscence With People With Dementia in Sub-Acute & Acute Care URL: http://medicine.unimelb.edu.au/__data/assets/pdf_file/0020/2471312/Using-reminiscence-with-people-with-dementia-in-acute-and-subacute-care-manual.pdf [accessed 2019-10-15]
25. Pishvaei M, Moghanloo RA, Moghanloo VA. The efficacy of treatment reminders of life with emphasis on integrative reminiscence on self-esteem and anxiety in widowed old men. Iran J Psychiatry 2015;10(1):19-24 [FREE Full text] [Medline: 26005476]
26. Moral JC, Terrero FB, Galán AS, Rodríguez TM. Effect of integrative reminiscence therapy on depression, well-being, integrity, self-esteem, and life satisfaction in older adults. J Posit Psychol 2014 Jul 18;10(3):240-247. [doi: 10.1080/17439760.2014.936968]
27. Woods B, O'Philbin L, Farrell EM, Spector AE, Orrell M. Reminiscence therapy for dementia. Cochrane Database Syst Rev 2018 Mar 1;3:CD001120 [FREE Full text] [doi: 10.1002/14651858.CD001120.pub3] [Medline: 29493789]
28. Chapoulie E, Guerchouche R, Petit P, Chaurasia G, Robert P, Drettakis G. Reminiscence Therapy using Image-Based Rendering in VR. In: Proceedings of the Conference on IEEE Virtual Reality. 2014 Presented at: VR'14; March 29-April 2, 2014; Minneapolis, MN, USA. [doi: 10.1109/vr.2014.6802049]
29. Lazar A, Thompson H, Demiris G. A systematic review of the use of technology for reminiscence therapy. Health Educ Behav 2014 Oct;41(1 Suppl):51S-61S [FREE Full text] [doi: 10.1177/1090198114537067] [Medline: 25274711]
30. Subramaniam P, Woods B. Digital life storybooks for people with dementia living in care homes: an evaluation. Clin Interv Aging 2016;11:1263-1276 [FREE Full text] [doi: 10.2147/CIA.S111097] [Medline: 27698556]
31. Ryan AA, McCauley CO, Laird EA, Gibson A, Mulvenna MD, Bond R, et al. 'There is still so much inside': the impact of personalised reminiscence, facilitated by a tablet device, on people living with mild to moderate dementia and their family carers. Dementia (London) 2020 May;19(4):1131-1150. [doi: 10.1177/1471301218795242] [Medline: 30193535]
32. Laird EA, Ryan A, McCauley C, Bond RB, Mulvenna MD, Curran KJ, et al. Using mobile technology to provide personalized reminiscence for people living with dementia and their carers: appraisal of outcomes from a quasi-experimental study. JMIR Ment Health 2018 Sep 11;5(3):e57 [FREE Full text] [doi: 10.2196/mental.9684] [Medline: 30206053]
33. Astell AJ, Ellis MP, Bernardi L, Alm N, Dye R, Gowans G, et al. Using a touch screen computer to support relationships between people with dementia and caregivers. Interact Comput 2010 Jul;22(4):267-275. [doi: 10.1016/j.intcom.2010.03.003]
34. Benoit M, Guerchouche R, Petit P, Chapoulie E, Manera V, Chaurasia G, et al. Is it possible to use highly realistic virtual reality in the elderly? A feasibility study with image-based rendering. Neuropsychiatr Dis Treat 2015;11:557-563 [FREE Full text] [doi: 10.2147/NDT.S73179] [Medline: 25834437]
35. Ma M, Zheng H. Virtual reality and serious games in healthcare. In: Brahnam S, Jain LC, editors. Advanced Computational Intelligence Paradigms in Healthcare 6: Virtual Reality in Psychotherapy, Rehabilitation, and Assessment. Berlin, Heidelberg: Springer; 2011:169-192.
36. Witmer BG, Jerome CJ, Singer MJ. The factor structure of the presence questionnaire. Presence 2005 Jun;14(3):298-312. [doi: 10.1162/105474605323384654]
37. Slobounov S, Ray W, Johnson B, Slobounov E, Newell K. Modulation of cortical activity in 2D versus 3D virtual reality environments: an EEG study. Int J Psychophysiol 2015 Mar;95(3):254-260. [doi: 10.1016/j.ijpsycho.2014.11.003] [Medline: 25448267]
38. Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, et al. Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]
39. Krijn M, Emmelkamp P, Olafsson R, Biemond R. Virtual reality exposure therapy of anxiety disorders: a review. Clin Psychol Rev 2004 Jul;24(3):259-281. [doi: 10.1016/j.cpr.2004.04.001] [Medline: 15245832]
40. Rose T, Nam C, Chen K. Immersion of virtual reality for rehabilitation - review. Appl Ergon 2018 May;69:153-161. [doi: 10.1016/j.apergo.2018.01.009] [Medline: 29477323]
41. Sankaranarayanan G, Li B, Manser K, Jones SB, Jones DB, Schwaitzberg S, et al. Face and construct validation of a next generation virtual reality (Gen2-VR) surgical simulator. Surg Endosc 2016 Mar;30(3):979-985 [FREE Full text] [doi: 10.1007/s00464-015-4278-7] [Medline: 26092010]
42. Reger GM, Holloway KM, Candy C, Rothbaum BO, Difede J, Rizzo AA, et al. Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. J Trauma Stress 2011 Feb;24(1):93-96. [doi: 10.1002/jts.20574] [Medline: 21294166]
43. Parsons T, Rizzo A. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 2008 Sep;39(3):250-261. [doi: 10.1016/j.jbtep.2007.07.007] [Medline: 17720136]
44. Hoffman HG, Patterson DR, Seibel E, Soltani M, Jewett-Leahy L, Sharar SR. Virtual reality pain control during burn wound debridement in the hydrotank. Clin J Pain 2008 May;24(4):299-304. [doi: 10.1097/AJP.0b013e318164d2cc] [Medline: 18427228]

45. Das DA, Grimmer KA, Sparnon AL, McRae SE, Thomas BH. The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: a randomized controlled trial [ISRCTN87413556]. BMC Pediatr 2005 Mar 3;5(1):1 [FREE Full text] [doi: 10.1186/1471-2431-5-1] [Medline: 15745448]
46. Huygelier H, Schraepen B, van Ee R, Abeele VV, Gillebert CR. Acceptance of immersive head-mounted virtual reality in older adults. Sci Rep 2019 Mar 14;9(1):4519 [FREE Full text] [doi: 10.1038/s41598-019-41200-6] [Medline: 30872760]
47. Rose V, Stewart I, Jenkins KG, Tabbaa L, Ang CS, Matsangidou M. Bringing the outside in: the feasibility of virtual reality with people with dementia in an inpatient psychiatric care setting. Dementia (London) 2019 Sep 11:- epub ahead of print. [doi: 10.1177/1471301219868036] [Medline: 31510801]
48. Roberts AR, de Schutter B, Franks K, Radina ME. Older adults' experiences with audiovisual virtual reality: perceived usefulness and other factors influencing technology acceptance. Clin Gerontol 2019;42(1):27-33. [doi: 10.1080/07317115.2018.1442380] [Medline: 29505343]
49. Brimelow R, Dawe B, Dissanayaka N. Preliminary research: virtual reality in residential aged care to reduce apathy and improve mood. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):165-170. [doi: 10.1089/cyber.2019.0286] [Medline: 31829729]
50. Optale G, Urgesi C, Busato V, Marin S, Piron L, Priftis K, et al. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. Neurorehabil Neural Repair 2010 May;24(4):348-357. [doi: 10.1177/1545968309353328] [Medline: 19934445]
51. Cho D, Lee S. Effects of virtual reality immersive training with computerized cognitive training on cognitive function and activities of daily living performance in patients with acute stage stroke: a preliminary randomized controlled trial. Medicine (Baltimore) 2019 Mar;98(11):e14752 [FREE Full text] [doi: 10.1097/MD.0000000000014752] [Medline: 30882644]
52. Micarelli A, Viziano A, Micarelli B, Augimeri I, Alessandrini M. Vestibular rehabilitation in older adults with and without mild cognitive impairment: effects of virtual reality using a head-mounted display. Arch Gerontol Geriatr 2019;83:246-256. [doi: 10.1016/j.archger.2019.05.008] [Medline: 31102927]
53. Marin RS. Differential diagnosis of apathy and related disorders of diminished motivation. Psychiatric Annals 1997;27(1):30-33. [doi: 10.3928/0048-5713-19970101-08]
54. McPherson S, Fairbanks L, Tiken S, Cummings JF, Back-Madruga C. Apathy and executive function in Alzheimer's disease. J Int Neuropsychol Soc 2002 Mar;8(3):373-381. [doi: 10.1017/s1355617702813182] [Medline: 11939696]
55. Tsoi T, Baillon S, Lindesay J. Early frontal executive impairment as a predictor of subsequent behavior disturbance in dementia. Am J Geriatr Psychiatry 2008 Feb;16(2):102-108. [doi: 10.1097/JGP.0b013e318151fb42] [Medline: 18239195]
56. Fishman KN, Ashbaugh AR, Lanctôt KL, Cayley ML, Herrmann N, Murray BJ, et al. Apathy, not depressive symptoms, as a predictor of semantic and phonemic fluency task performance in stroke and transient ischemic attack. J Clin Exp Neuropsychol 2018 Jun;40(5):449-461. [doi: 10.1080/13803395.2017.1371282] [Medline: 28911268]
57. Grossi D, Santangelo G, Barbarulo AM, Vitale C, Castaldo G, Proto MG, et al. Apathy and related executive syndromes in dementia associated with Parkinson's disease and in Alzheimer's disease. Behav Neurol 2013 Jan 1;27(4):515-522 [FREE Full text] [doi: 10.3233/BEN-129023] [Medline: 23242363]
58. Meyer A, Zimmermann R, Gschwandtner U, Hatz F, Bousleiman H, Schwarz N, et al. Apathy in Parkinson's disease is related to executive function, gender and age but not to depression. Front Aging Neurosci 2014;6:350 [FREE Full text] [doi: 10.3389/fnagi.2014.00350] [Medline: 25642187]
59. Shao Z, Janse E, Visser K, Meyer AS. What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. Front Psychol 2014;5:772 [FREE Full text] [doi: 10.3389/fpsyg.2014.00772] [Medline: 25101034]
60. Rosser A, Hodges JR. Initial letter and semantic category fluency in Alzheimer's disease, Huntington's disease, and progressive supranuclear palsy. J Neurol Neurosurg Psychiatry 1994 Nov;57(11):1389-1394 [FREE Full text] [doi: 10.1136/jnnp.57.11.1389] [Medline: 7964817]
61. Robert P, Lanctôt KL, Agüera-Ortiz L, Aalten P, Bremond F, Defrancesco M, et al. Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. Eur Psychiatry 2018 Oct;54:71-76 [FREE Full text] [doi: 10.1016/j.eurpsy.2018.07.008] [Medline: 30125783]
62. Radakovic R, Abrahams S. Developing a new apathy measurement scale: dimensional apathy scale. Psychiatry Res 2014 Nov 30;219(3):658-663. [doi: 10.1016/j.psychres.2014.06.010] [Medline: 24972546]
63. Manera V, Chapoulie E, Bourgeois J, Guerchouche R, David R, Ondrej J, et al. A feasibility study with image-based rendered virtual reality in patients with mild cognitive impairment and dementia. PLoS One 2016;11(3):e0151487 [FREE Full text] [doi: 10.1371/journal.pone.0151487] [Medline: 26990298]
64. Kim A, Darakjian N, Finley JM. Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. J Neuroeng Rehabil 2017 Feb 21;14(1):16 [FREE Full text] [doi: 10.1186/s12984-017-0225-2] [Medline: 28222783]
65. Parijat P, Lockhart TE. Can virtual reality be used as a gait training tool for older adults? Proc Hum Factors Ergon Soc 2011 Sep 1;55(1):157-161. [doi: 10.1177/1071181311551033]
66. Kang YJ, Ku J, Han K, Kim SI, Yu TW, Lee JH, et al. Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: preliminary study. Cyberpsychol Behav 2008 Jun;11(3):329-339. [doi: 10.1089/cpb.2007.0116] [Medline: 18537503]

67. Aged Care Virtual Reality. 2018. VR Programs for Health Benefits URL: http://www.agedcarevirtualreality.com/ [accessed 2019-08-20]
68. Reminiscence. 2018. About Us URL: https://www.reminiscience.co.uk/about-us [accessed 2019-08-20]
69. Rendever. 2018. Virtual Reality for Seniors URL: http://rendever.com/ [accessed 2019-08-20]
70. Saredakis D, Szpak A, Birckhead B, Keage HAD, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. Front Hum Neurosci 2020;14:96 [FREE Full text] [doi: 10.3389/fnhum.2020.00096] [Medline: 32300295]
71. Jorm AF, Mackinnon AJ, Henderson AS, Scott R, Christensen H, Korten AE, et al. The psychogeriatric assessment scales: a multi-dimensional alternative to categorical diagnoses of dementia and depression in the elderly. Psychol Med 1995 May;25(3):447-460. [doi: 10.1017/s0033291700033377] [Medline: 7480426]
72. Clarke DE, Reekum RV, Simard M, Streiner DL, Freedman M, Conn D. Apathy in dementia: an examination of the psychometric properties of the apathy evaluation scale. J Neuropsychiatry Clin Neurosci 2007;19(1):57-64. [doi: 10.1176/jnp.2007.19.1.57] [Medline: 17308228]
73. Oculus. 2019. Oculus Go URL: https://www.oculus.com/go/ [accessed 2019-07-10]
74. Jorm A, Mackinnon A. Australian Government Department of Health. 1995. The Psychogeriatric Assessment Scales URL: https://www.health.gov.au/sites/default/files/documents/2019/12/psychogeriatric-assessment-scales-pas-user-guide.pdf [accessed 2019-08-20]
75. Dyer SM, Gnanamanickam ES, Liu E, Whitehead C, Crotty M. Diagnosis of dementia in residential aged care settings in Australia: an opportunity for improvements in quality of care? Australas J Ageing 2018 Dec;37(4):E155-E158 [FREE Full text] [doi: 10.1111/ajag.12580] [Medline: 30188008]
76. Jorm A, Christensen H, Jacomb PA, Korten AE, Mackinnon AJ. The cognitive decline scale of the psychogeriatric assessment scales (PAS): longitudinal data on its validity. Int J Geriatr Psychiatry 2001 Mar;16(3):261-265. [doi: 10.1002/gps.326] [Medline: 11288159]
77. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
78. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. Virtual Reality 2016 Apr 26;20(2):101-125. [doi: 10.1007/s10055-016-0285-9]
79. Bouchard S, Robillard G, Renaud P. Revising the factor structure of the simulator sickness questionnaire. Annu Rev CyberTherapy Telemed 2007;5:128-137 [FREE Full text]
80. Usoh M, Catena E, Arman S, Slater M. Using presence questionnaires in reality. Presence 2000 Oct;9(5):497-503. [doi: 10.1162/105474600566989]
81. Schwind V, Knierim P, Haas N, Henze N. Using Presence Questionnaires in Virtual Reality. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300590]
82. Spreen O, Strauss E. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Second Edition. New York, NY: Oxford University Press; 1998.
83. Merhi O, Faugloire E, Flanagan M, Stoffregen TA. Motion sickness, console video games, and head-mounted displays. Hum Factors 2007 Oct;49(5):920-934. [doi: 10.1518/001872007X230262] [Medline: 17915607]
84. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull 1980;87(2):245-251. [doi: 10.1037/0033-2909.87.2.245]
85. Kennedy R, Drexler J, Compton D, Stanney K, Lanham D, Harm D. Congural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences. In: Hettinger LJ, Haas MW, editors. Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues. Mahwah, NJ: Lawrence Erlbaum Associates; 2003:247-278.
86. Ellis JM, Doyle CJ, Selvarajah S. The relationship between apathy and participation in therapeutic activities in nursing home residents with dementia: evidence for an association and directions for further research. Dementia (London) 2016 Jul;15(4):494-509. [doi: 10.1177/1471301214527300] [Medline: 24670286]
87. Subramaniam P, Woods B. Towards the therapeutic use of information and communication technology in reminiscence work for people with dementia: a systematic review. Int J Comput Healthcare 2010;1(2):106. [doi: 10.1504/ijcih.2010.037457]
88. Bermingham A, O'Rourke J, Gurrin C, Collins R, Irving K, Smeaton A. Automatically Recommending Multimedia Content for Use in Group Reminiscence Therapy. In: Proceedings of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare. 2013 Presented at: MIIRH'13; October 22, 2013; Barcelona, Spain. [doi: 10.1145/2505323.2505333]
89. Pinquart M, Forstmeier S. Effects of reminiscence interventions on psychosocial outcomes: a meta-analysis. Aging Ment Health 2012;16(5):541-558. [doi: 10.1080/13607863.2011.651434] [Medline: 22304736]
90. Elias SS, Neville C, Scott T. The effectiveness of group reminiscence therapy for loneliness, anxiety and depression in older adults in long-term care: a systematic review. Geriatr Nurs 2015;36(5):372-380. [doi: 10.1016/j.gerinurse.2015.05.004] [Medline: 26099638]

91. Yasuda K, Kuwabara K, Kuwahara N, Abe S, Tetsutani N. Effectiveness of personalised reminiscence photo videos for individuals with dementia. Neuropsychol Rehabil 2009 Aug;19(4):603-619. [doi: 10.1080/09602010802586216] [Medline: 19235628]
92. Subramaniam P, Woods B. The impact of individual reminiscence therapy for people with dementia: systematic review. Expert Rev Neurother 2012 May;12(5):545-555. [doi: 10.1586/ern.12.35] [Medline: 22550983]
93. Baker S, Waycott J, Robertson E, Carrasco R, Neves B, Hampson R, et al. Evaluating the use of interactive virtual reality technology with older adults living in residential aged care. Inf Process Manag 2020 May;57(3):102105 [FREE Full text] [doi: 10.1016/j.ipm.2019.102105]
94. Vaughan RM, Coen RF, Kenny R, Lawlor BA. Preservation of the semantic verbal fluency advantage in a large population-based sample: normative data from the TILDA study. J Int Neuropsychol Soc 2016 May;22(5):570-576. [doi: 10.1017/S1355617716000291] [Medline: 27055803]
95. Tombaugh T, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Arch Clin Neuropsychol 1999 Feb;14(2):167-177. [Medline: 14590600]
96. Okumura Y, Tanimukai S, Asada T. Effects of short-term reminiscence therapy on elderly with dementia: a comparison with everyday conversation approaches. Psychogeriatrics 2008;8(3):124-133. [doi: 10.1111/j.1479-8301.2008.00236.x]
97. Drijgers RL, Verhey FR, Leentjens AF, Köhler S, Aalten P. Neuropsychological correlates of apathy in mild cognitive impairment and Alzheimer's disease: the role of executive functioning. Int Psychogeriatr 2011 Oct;23(8):1327-1333. [doi: 10.1017/S1041610211001037] [Medline: 21729413]
98. Massimi M, Berry E, Browne G, Smyth G, Watson P, Baecker RM. An exploratory case study of the impact of ambient biographical displays on identity in a patient with Alzheimer's disease. Neuropsychol Rehabil 2008;18(5-6):742-765. [doi: 10.1080/09602010802130924] [Medline: 18609021]
99. Allan L, Ballard C, Burn D, Kenny R. Prevalence and severity of gait disorders in Alzheimer's and non-Alzheimer's dementias. J Am Geriatr Soc 2005 Oct;53(10):1681-1687. [doi: 10.1111/j.1532-5415.2005.53552.x] [Medline: 16181166]
100. DeLuca SA, Korcuska LA, Oberstar BH, Rosenthal ML, Welsh PA, Topol EJ. Are we promoting true informed consent in cardiovascular clinical trials? J Cardiovasc Nurs 1995 Apr;9(3):54-61. [doi: 10.1097/00005082-199504000-00007] [Medline: 7782813]
101. McMurdo M, Roberts H, Parker S, Wyatt N, May H, Goodman C, AgeAgeing Specialty Group, NIHR, Comprehensive Clinical Research Network. Improving recruitment of older people to research through good practice. Age Ageing 2011 Nov;40(6):659-665. [doi: 10.1093/ageing/afr115] [Medline: 21911335]

Abbreviations

AES: Apathy Evaluation Scale
HMD: head-mounted display
PAS: Psychogeriatric Assessment Scale
PTSD: post-traumatic stress disorder
SSQ: Simulator Sickness Questionnaire
SUS: Slater-Usoh-Steed Presence Questionnaire
VR: virtual reality

Edited by G Eysenbach; submitted 29.12.19; peer-reviewed by H Huygelier, L Campbell; comments to author 10.02.20; revised version received 26.03.20; accepted 08.04.20; published 26.06.20.

Please cite as:
Saredakis D, Keage HAD, Corlis M, Loetscher T
Using Virtual Reality to Improve Apathy in Residential Aged Care: Mixed Methods Study
J Med Internet Res 2020;22(6):e17632
URL: http://www.jmir.org/2020/6/e17632/
doi:10.2196/17632
PMID:32469314

Short Paper

Evaluating the Need to Address Digital Literacy Among Hospitalized Patients: Cross-Sectional Observational Study

Hanna Vollbrecht[1], BS; Vineet Arora[2], MD; Sebastian Otero[2], BA; Kyle Carey[3], MPH; David Meltzer[3], MD, PhD; Valerie G Press[2], MD, MPH

[1]Pritzker School of Medicine, University of Chicago, Chicago, IL, United States

[2]Section of General Internal Medicine, Department of Medicine, University of Chicago, Chicago, IL, United States

[3]Section of Hospital Medicine, Department of Medicine, University of Chicago, Chicago, IL, United States

Corresponding Author:
Valerie G Press, MD, MPH
Section of General Internal Medicine
Department of Medicine
University of Chicago
5841 S Maryland Avenue
MC2007
Chicago, IL, 60637
United States
Phone: 1 773 702 5170
Email: vpress@bsd.uchicago.edu

Abstract

Background: Technology is a potentially powerful tool to assist patients with transitions of care during and after hospitalization. Patients with low health literacy who are predisposed to poor health outcomes are particularly poised to benefit from such interventions. However, this population may lack the ability to effectively engage with technology. Although prior research studied the role of health literacy in technology access/use among outpatients, hospitalized patient populations have not been investigated in this context. Further, with the rapid uptake of technology, access may no longer be pertinent, and differences in technological capabilities may drive the current digital divide. Thus, characterizing the digital literacy of hospitalized patients across health literacy levels is paramount.

Objective: We sought to determine the relationship between health literacy level and technological access, use, and capability among hospitalized patients.

Methods: Adult inpatients completed a technology survey that asked about technology access/use and online capabilities as part of an ongoing quality of care study. Participants' health literacy level was assessed utilizing the 3-question Brief Health Literacy Screen. Descriptive statistics, bivariate chi-squared analyses, and multivariate logistic regression analyses (adjusting for age, race, gender, and education level) were performed. Using Bonferroni correction for the 18 tests, the threshold *P* value for significance was <.003.

Results: Among 502 enrolled participants, the mean age was 51 years, 71.3% (358/502) were African American, half (265/502, 52.8%) were female, and half (253/502, 50.4%) had at least some college education. Over one-third (191/502, 38.0%) of participants had low health literacy. The majority of participants owned devices (owned a smartphone: 116/173, 67.1% low health literacy versus 235/300, 78.3% adequate health literacy, *P*=.007) and had used the Internet previously (143/189, 75.7% low health literacy versus 281/309, 90.9% adequate health literacy, *P*<.001). Participants with low health literacy were more likely to report needing help performing online tasks (133/189, 70.4% low health literacy versus 135/303, 44.6% adequate health literacy, *P*<.001). In the multivariate analysis, when adjusting for age, race, gender, and education level, we found that low health literacy was not significantly associated with a lower likelihood of owning smartphones (OR: 0.8, 95% CI 0.5-1.4; *P*=.52) or using the internet ever (OR: 0.5, 95% CI 0.2-0.9; *P*=.02). However, low health literacy remained significantly associated with a higher likelihood of needing help performing any online task (OR: 2.2, 95% CI 1.3-3.6; *P*=.002).

Conclusions: The majority of participants with low health literacy had access to technological devices and had used the internet previously, but they were unable to perform online tasks without assistance. The barriers patients face in using online health information and other health information technology may be more related to online capabilities rather than to technology access.

When designing and implementing technological tools for hospitalized patients, it is important to ensure that patients across digital literacy levels can both understand and use them.

(J Med Internet Res 2020;22(6):e17519) doi:10.2196/17519

KEYWORDS

health literacy; digital literacy; hospitalization; technology

Introduction

Technology-based interventions have the potential to improve care transitions; however, they can also exacerbate existing health disparities. For instance, hospitalization represents a time of vulnerable care transitions [1] when technology-based interventions could improve patient engagement and outcomes. Examples include patient portals [2], educational videos [3,4], mobile apps [5], telehealth [6], and remote monitoring [7], which have the possibility of improving patient-provider communication and patient education at the time of hospitalization and discharge, when patients may be coping with a new diagnosis or needing assistance with controlling a chronic disease. However, these interventions can only be broadly effective if all patients are able to access, use and effectively understand them. Historically, the digital divide concept was characterized by differences in access to technology and largely driven by socioeconomic status (SES), age, and race [8]. Currently, with increasing access to technology-based devices, a shift to a digital capability divide may be even more salient [9].

Prior research has shown that health literacy is an important contributor to the digital capability divide, with patients with lower health literacy being 7%-47% less likely to access and have the ability to use technology [10,11]. At the same time, patients with low health literacy have worse health outcomes [12], increased risk of poor vision [13], longer hospital stays [14], increased hospital-to-home transitional care needs [15], and increased readmission risk, especially among older patients [16]. Therefore, technology-based interventions could be a mechanism to improve long-term outcomes for hospitalized patients with low health literacy, while adding complexity for some patients. Research investigating the relationship between health literacy and technology has been primarily conducted among community-dwelling and outpatient populations but has not been evaluated among hospitalized patients.

The relationship between health literacy and technology use may be particularly relevant for hospitalized patients, since many care transition interventions rely on technology. Prior research has also suggested that health literacy is dynamic, with hospitalization representing a period when health literacy may acutely decrease [17]. It is also very likely that technology access and capabilities are not static, but dynamic. Hospitalization provides an important assessment time point to understand if and how technology can be utilized to improve health. Finally, hospitalized patients likely differ from those in community-dwelling or outpatient settings, making them an important population to characterize. Therefore, we sought to determine the relationship between health literacy level and technology access, use, and digital capabilities among hospitalized adult general medicine patients.

Methods

We performed a cross-sectional observational study among adult inpatients at the University of Chicago Medicine as part of a large, ongoing study of inpatient quality of care [18]. Inclusion criteria included being hospitalized on a general medicine service, 18 years or older, and English-speaking. Exclusion criteria included an inability to provide consent and prior participation in the study. The University of Chicago Biological Sciences Division Institutional Review Board approved this protocol (#IRB16-0763).

We used a survey that was comprised primarily of national benchmarked Pew Research Center survey questions [19] that were categorized into 3 domains of technology: access, use, and capabilities. All variables in these domains were binary (yes/no). To assess technology access, participants were asked whether they owned a smartphone, computer, or tablet; whether they had a texting plan; and whether they had Wi-Fi at home. To assess technology use, participants were asked if they used the internet and whether they used the internet for health-related reasons. To assess digital capabilities, participants were asked if they knew how to perform a given online task or if they would need help. A summary measure was constructed for needing help with any online task, in which participants were categorized as needing help if they responded yes to needing help with one or more tasks. Participants were assigned as having either low or adequate health literacy based on the 3-question Brief Health Literacy Screen (BHLS). Each item in the BHLS is scored on a Likert scale from 0 to 4, with a score of 2 or less on any item identifying a participant as having low health literacy [20]. The BHLS and technology survey were orally administered.

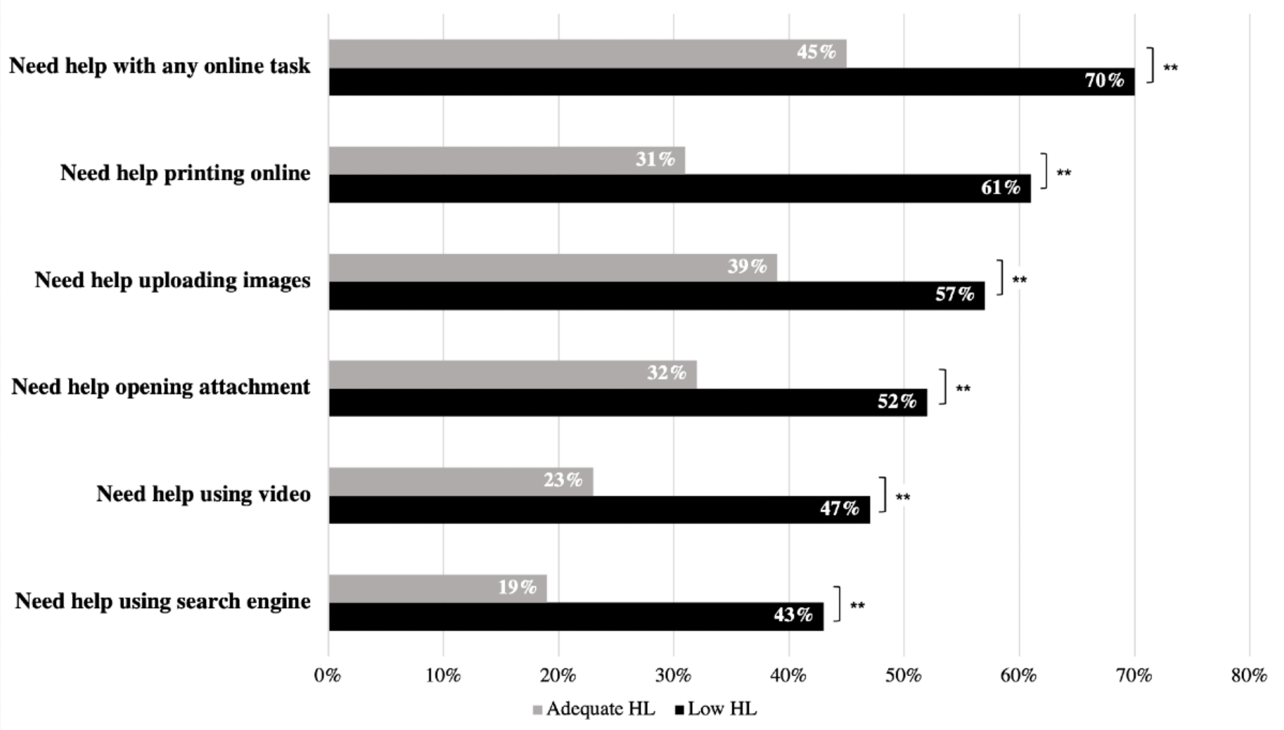
Differences in technology access, use, and capabilities by health literacy level were analyzed using chi-squared tests. Multivariate logistic regression analyses were conducted to determine differences in technology access, use, and capabilities adjusting for health literacy (binary), age (continuous), gender (binary), race (white, non-Hispanic black, other), and education (less than any college versus some college or more). All analyses were performed using STATA version 15.0 (StataCorp), with $P<.003$ defining statistical significance based on Bonferroni correction.

Results

From January 30, 2014 through May 10, 2018, 502 participants were enrolled in the study and completed the survey. Of these 502 participants, the mean age was 51 years, 358 (71.3%) were Black (non-Hispanic), 265 (52.8%) were female, and 253

(50.4%) had at least some college education. Out of 502 participants, 191 (38.0%) had low health literacy. Compared to participants with adequate health literacy, participants with low health literacy were less likely to own a desktop (49/191 [25.7%] low health literacy versus 130/311 [41.8%] adequate health literacy, *P*<.001) or laptop (63/191 [33.0%] versus 173/311 [55.6%], *P*<.001). In contrast, there was no significant difference in ownership of tablets (64/191 [33.5%] versus 139/311 [44.7%], *P*=.01) or smartphones (116/173 [67.0%] versus 235/300 [78.3%], *P*=.007) by health literacy level. Participants with low health literacy were less likely to report using the internet ever (143/189 [75.7%] versus 281/309 [90.9%], *P*<.001), daily internet use (91/189 [48.1%] versus 210/309 [68.0%], *P*<.001), or searching for health information online (95/155 [61.3%] versus 222/288 [77.1%], *P*<.001). Participants with low health literacy were more likely to report needing help for all online tasks queried (Figure 1).

Figure 1. Percent of participants that report needing help with online tasks by health literacy (HL) level. ** demonstrates significance of *P*<.003 for Bonferroni correction. Raw numbers of participants needing help with each task are as follows: Need help with any online task – adequate HL 135/303 (44.6%), low HL 133/189 (70.4%); Need help printing online – adequate HL 95/303 (31.4%), low HL 116/189 (61.4%); Need help uploading images – adequate HL 119/303 (39.3%), low HL 107/189 (56.6%); Need help opening attachment – adequate HL 96/303 (31.7%), low HL 99/189 (52.4%); Need help using video – adequate HL 61/257 (23.7%), low HL 81/172 (47.1%); Need help using search engine – adequate HL 59/303 (19.5%), low HL 82/189 (43.4%).



In multivariate analysis, we found that health literacy remained significantly associated with some aspects of technology access/use, but not others (Table 1). For example, low health literacy was still associated with a lower likelihood of owning a laptop (*P*<.001), but not with owning a smartphone (*P*=.50). Low health literacy level was not significantly associated with a lower likelihood of internet use (*P*=.02) or searching for health information online (*P*=.10). Low health literacy was associated with a higher likelihood of needing help with online tasks overall. Further results from multivariate logistic regression analyses are provided in Multimedia Appendices 1 and 2.

Table 1. The relationship of low health literacy and technology access, use, and capabilities[a].

Low health literacy[b]	OR[c] (95% CI)	*P* value	AOR[d] (95% CI)	*P* value
Technology access				
Own desktop	0.5 (0.3-0.7)	<.001	0.6 (0.4-0.98)	.04
Own laptop	0.4 (0.3-0.6)	<.001	0.5 (0.3-0.8)	.001
Own tablet	0.6 (0.4-0.9)	.01	0.8 (0.5-1.2)	.33
Own smartphone	0.6 (0.4-0.9)	.01	0.8 (0.5-1.4)	.52
Wi-Fi at home	0.4 (0.3-0.7)	<.001	0.7 (0.4-1.0)	.07
Text messaging plan-any	0.3 (0.2-0.6)	<.001	0.3 (0.2-0.7)	.002
Unlimited text plan	0.5 (0.3-0.8)	.007	0.6 (0.3-1.0)	.05
Technology use				
Ever internet use	0.3 (0.2-0.5)	<.001	0.5 (0.2-0.9)	.02
Daily internet use	0.4 (0.3-0.6)	<.001	0.6 (0.4-0.97)	.04
Search health info online	0.5 (0.3-0.7)	<.001	0.7 (0.4-1.1)	.11
Post health info online	0.6 (0.3-0.997)	.05	0.6 (0.4-1.0)	.07
Download app	0.6 (0.3-0.9)	.02	0.8 (0.5-1.4)	.40
Technology capabilities				
Need help with any online task[e]	3.0 (2-4.3)	<.001	2.2 (1.3-3.6)	.002
Need help to print online materials	3.5 (2.4-5.1)	<.001	2.7 (1.7-4.4)	<.001
Need help to upload images	2.0 (1.4-2.9)	<.001	1.4 (0.9-2.2)	.20
Need help to open attachment	2.4 (1.6-3.4)	<.001	1.7 (1.1-2.8)	.03
Need help to use video	2.9 (1.9-4.3)	<.001	2.5 (1.4-4.2)	.001
Need help to use search engine	3.2 (2.1-4.7)	<.001	2.1 (1.3-3.5)	.003

[a] Using Bonferroni correction for the 18 tests, the *P* value threshold for significance is <.003.

[b] Low health literacy is a binary variable where low HL = 0 denotes participants with adequate health literacy and low HL = 1 denotes participants with low health literacy.

[c] OR: unadjusted odds ratio.

[d] AOR: best fit adjusted odds ratio (adjusted for age [continuous], gender [binary], race [white, black, other], and education [less than any college versus some college or more]).

[e] Compilation of all items beginning with "Need help…", with participants categorized as needing help if they responded Yes to needing help with 1 or more of the online tasks listed.

Discussion

We found that despite most participants having access to smartphones and using the internet, hospitalized patients with low health literacy were significantly more likely to need help with online tasks. This raises the question of whether inpatients with low health literacy are able to effectively utilize technology, even if they have access to do so. Although our study suggests that certain devices, such as smartphones, may have appealing interfaces for patients across literacy levels, it also underlines that access is not synonymous with ability. This is consistent with a recent study showing disparities in hospitalized patients' interest in patient portals [21]. Patients who were older, African-American, non-English speaking, or homeless were less likely to want to use patient portals and the second most commonly cited reason for this was an inability to use the internet. If technology is to be implemented more broadly during and after hospitalization for patient use, we must do so in a way that is palatable, engaging, feasible, and equitable for diverse populations. It is not enough to design these resources in a health-literate manner; we must also ensure that patients have the necessary digital skills to utilize them. Interestingly, results from multivariate analyses suggested that in the cases where health literacy is no longer associated with certain technological capabilities (specifically, using a search engine and uploading images), age and education are significantly associated. This suggests that in years to come, as more people are raised with technology, the digital divide will likely shrink. Examining trends of digital capabilities over the period of our study would be interesting future work.

Some hospitals are measuring health literacy among their patients to identify those who may be at high risk for nonunderstanding and poorer health outcomes [22]. In addition, our findings suggest that it may be important to assess digital literacy if hospitals are promoting the use of technology for patients' self-care. It is also possible that a universal approach to digital literacy could be useful. Health literacy universal precautions have been proposed; they encourage providers and health care systems to approach all patients with the assumption that they may not understand health information. This is based on the idea that health literacy is a dynamic process that is not only determined by individuals' abilities but also the complexities of the system [17]. In the same manner, hospitals could approach all patients with universal digital literacy precautions, ensuring that interventions are designed to be accessible and usable for patients across literacy levels and that patients have the ability to utilize technology as part of implementation. Although this requires up-front increased personnel to educate patients, long-term, technological resources could be cost-saving [23].

Recently, electronic health (eHealth) literacy has gained traction as an important aspect of patients' ability to obtain and use information from online sources. eHealth literacy represents a complex interplay between multiple literacies, including health literacy and digital literacy [24], which we measured in this study. However, these two literacies alone likely do not adequately encompass all the skills necessary to effectively utilize technology-based health resources. Future studies could use eHealth literacy assessment tools to determine if and when eHealth literacy should be assessed prior to prescribing technology-based interventions in the hospital setting. Additionally, a recent study found that health literacy and eHealth literacy were not significantly correlated [25], suggesting that both may need to be studied to determine patients' abilities to use technology-based health resources. Future studies could aim to examine eHealth literacy among hospitalized patients and whether health literacy, digital literacy, and eHealth literacy are correlated.

Limitations of our study include being a single-site study and using self-reported measures of technology access, use, capabilities, and health literacy. Further, our study did not measure eHealth literacy directly. Future studies could focus specifically on the eHealth literacy of hospitalized patients and how it relates to their ability to use health technologies during and after hospitalization.

In summary, our results indicated that health literacy is an important contributor to the digital capability divide among hospitalized patients and helped to identify areas of future research. Technology provides both opportunities for improvement during care transitions and potential pitfalls. To mitigate these pitfalls, health literacy and digital literacy should be accounted for when considering how best to implement technology-based interventions across hospital-to-home transitions of care.

Acknowledgments

We would like to acknowledge Mary Akel, Ashley Hull, Eddie Kim, John Kim, Daniel Chang, and Brenda Munguia for their assistance with this project. This project was funded in part by an NIH NHLBI K23 (HL 118151), an American Lung Foundation Social Behavioral Award, and a Chicago Center for Diabetes Translation Research Pilot and Feasibility Grant. An NIH NIA grant #4T35AG029795-10 provided support to HV. VA receives funding from NHLBI K24HL136859.

Conflicts of Interest

VP reports previously consulting for Round Glass and currently serving as a subject matter expert for Vizient and as a consultant for Humana.

Multimedia Appendix 1

Results of multivariate logistic regression analyses in which low health literacy (HL) adjusted odds ratio (AOR) was significant.
[DOCX File , 17 KB - jmir_v22i6e17519_app1.docx]

Multimedia Appendix 2

Results of multivariate logistic regression analyses in which low health literacy (HL) adjusted odds ratio (AOR) was not significant.
[DOCX File , 26 KB - jmir_v22i6e17519_app2.docx]

References

1. Kripalani S, Jackson AT, Schnipper JL, Coleman EA. Promoting effective transitions of care at hospital discharge: a review of key issues for hospitalists. J Hosp Med 2007 Sep;2(5):314-323. [doi: 10.1002/jhm.228] [Medline: 17935242]
2. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient Portals Facilitating Engagement With Inpatient Electronic Medical Records: A Systematic Review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
3. Press VG, Kelly CA, Kim JJ, White SR, Meltzer DO, Arora VM. Virtual Teach-To-Goal™ Adaptive Learning of Inhaler Technique for Inpatients with Asthma or COPD. J Allergy Clin Immunol Pract 2017;5(4):1032-1039.e1 [FREE Full text] [doi: 10.1016/j.jaip.2016.11.018] [Medline: 28065689]
4. Press VG, Arora VM, Kelly CA, Carey KA, White SR, Wan W. Effectiveness of Virtual vs In-Person Inhaler Education for Hospitalized Patients With Obstructive Lung Disease: A Randomized Clinical Trial. JAMA Netw Open 2020 Jan 03;3(1):e1918205 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18205] [Medline: 31899529]

5. Habib B, Fosser SM, Buckeridge DL, Weir DL, Bustillo M, Thakur M, et al. Evaluation of a Mobile Application to Enhance Medication Management Following Hospital Discharge: Study Protocol for a Pilot Randomized Controlled Trial. Stud Health Technol Inform 2019 Aug 21;264:1929-1930. [doi: 10.3233/SHTI190718] [Medline: 31438412]
6. O'Connor M, Asdornwised U, Dempsey ML, Huffenberger A, Jost S, Flynn D, et al. Using Telehealth to Reduce All-Cause 30-Day Hospital Readmissions among Heart Failure Patients Receiving Skilled Home Health Services. Appl Clin Inform 2016;7(2):238-247 [FREE Full text] [doi: 10.4338/ACI-2015-11-SOA-0157] [Medline: 27437037]
7. Hale TM, Jethwani K, Kandola MS, Saldana F, Kvedar JC. A Remote Medication Monitoring System for Chronic Heart Failure Patients to Reduce Readmissions: A Two-Arm Randomized Pilot Study. J Med Internet Res 2016 Apr 17;18(5):e91 [FREE Full text] [doi: 10.2196/jmir.5256] [Medline: 27154462]
8. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the Internet, and the digital divide. Health Aff (Millwood) 2000;19(6):255-265. [doi: 10.1377/hlthaff.19.6.255] [Medline: 11192412]
9. Wei K, Teo H, Chan HC, Tan BCY. Conceptualizing and Testing a Social Cognitive Model of the Digital Divide. Information Systems Research 2011 Mar;22(1):170-187. [doi: 10.1287/isre.1090.0273]
10. Bailey SC, O'Conor R, Bojarski EA, Mullen R, Patzer RE, Vicencio D, et al. Literacy disparities in patient access and health-related use of Internet and mobile technologies. Health Expect 2015 Dec;18(6):3079-3087 [FREE Full text] [doi: 10.1111/hex.12294] [Medline: 25363660]
11. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017 Apr 01;24(e1):e47-e54 [FREE Full text] [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
12. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
13. Press VG, Shapiro MI, Mayo AM, Meltzer DO, Arora VM. More than meets the eye: relationship between low health literacy and poor vision in hospitalized patients. J Health Commun 2013;18 Suppl 1:197-204 [FREE Full text] [doi: 10.1080/10810730.2013.830346] [Medline: 24093356]
14. Jaffee EG, Arora VM, Matthiesen MI, Meltzer DO, Press VG. Health Literacy and Hospital Length of Stay: An Inpatient Cohort Study. J Hosp Med 2017 Dec;12(12):969-973. [doi: 10.12788/jhm.2848] [Medline: 29236095]
15. Boyle J, Speroff T, Worley K, Cao A, Goggins K, Dittus RS, et al. Low Health Literacy Is Associated with Increased Transitional Care Needs in Hospitalized Patients. J Hosp Med 2017 Nov;12(11):918-924. [doi: 10.12788/jhm.2841] [Medline: 29091980]
16. Jaffee EG, Arora VM, Matthiesen MI, Hariprasad SM, Meltzer DO, Press VG. Postdischarge Falls and Readmissions: Associations with Insufficient Vision and Low Health Literacy among Hospitalized Seniors. J Health Commun 2016;21(sup2):135-140. [doi: 10.1080/10810730.2016.1179371] [Medline: 27660917]
17. Koh HK, Brach C, Harris LM, Parchman ML. A proposed 'health literate care model' would constitute a systems approach to improving patients' engagement in care. Health Aff (Millwood) 2013 Feb;32(2):357-367. [doi: 10.1377/hlthaff.2012.1205] [Medline: 23381529]
18. Meltzer D, Manning WG, Morrison J, Shah MN, Jin L, Guth T, et al. Effects of physician experience on costs and outcomes on an academic general medicine service: results of a trial of hospitalists. Ann Intern Med 2002 Dec 3;137(11):866-874. [Medline: 12458986]
19. Pew Internet Project. Datasets, Pew Research Center: Internet, Science & Tech. 2016. URL: http://www.pewinternet.org/datasets/ [accessed 2019-11-01]
20. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008 May;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]
21. Sadasivaiah S, Lyles CR, Kiyoi S, Wong P, Ratanawongsa N. Disparities in Patient-Reported Interest in Web-Based Patient Portals: Survey at an Urban Academic Safety-Net Hospital. J Med Internet Res 2019 Mar 26;21(3):e11421 [FREE Full text] [doi: 10.2196/11421] [Medline: 30912747]
22. Cawthon C, Mion LC, Willens DE, Roumie CL, Kripalani S. Implementing routine health literacy assessment in hospital and primary care patients. Jt Comm J Qual Patient Saf 2014 Feb;40(2):68-76 [FREE Full text] [doi: 10.1016/s1553-7250(14)40008-4] [Medline: 24716329]
23. Sanyal C, Stolee P, Juzwishin D, Husereau D. Economic evaluations of eHealth technologies: A systematic review. PLoS One 2018;13(6):e0198112 [FREE Full text] [doi: 10.1371/journal.pone.0198112] [Medline: 29897921]
24. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
25. Monkman H, Kushniruk AW, Barnett J, Borycki EM, Greiner LE, Sheets D. Are Health Literacy and eHealth Literacy the Same or Different? Stud Health Technol Inform 2017;245:178-182. [Medline: 29295077]

Abbreviations

BHLS: Brief Health Literacy Screen
eHealth: electronic health
HL: health literacy
SES: socioeconomic status

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by B Halpenny, Z Ma, C Geukes; comments to author 10.02.20; revised version received 19.03.20; accepted 21.03.20; published 04.06.20.

Please cite as:
Vollbrecht H, Arora V, Otero S, Carey K, Meltzer D, Press VG
Evaluating the Need to Address Digital Literacy Among Hospitalized Patients: Cross-Sectional Observational Study
J Med Internet Res 2020;22(6):e17519
URL: https://www.jmir.org/2020/6/e17519
doi:10.2196/17519
PMID:32496196

Original Paper

Internet Use, eHealth Literacy, and Dietary Supplement Use Among Young Adults in Pakistan: Cross-Sectional Study

Amina Tariq[1], PhD; Shanchita R Khan[1], PhD; Amna Basharat[2], PhD

[1]School of Public Health and Social Work, Faculty of Health, Queensland University of Technology, Kelvin Grove, Australia

[2]Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan

Corresponding Author:
Amina Tariq, PhD
School of Public Health and Social Work
Faculty of Health
Queensland University of Technology
O Block, D Wing
Victoria Park Road
Kelvin Grove, QLD 4059
Australia
Phone: 61 7 3138 5612
Email: a.tariq@qut.edu.au

Abstract

Background: Increased access to the internet has facilitated widespread availability of health information. Thus, electronic health (eHealth) literacy—the ability to seek, find, understand, and appraise health information from electronic resources and apply that knowledge in making a health-related decision—is a crucial skill. Despite the increasing use of the internet as a source of health information in developing countries, only a few studies have examined the eHealth literacy of young adults, who frequently use the internet to access health information in these developing countries.

Objective: The aim of this study was to assess the patterns of internet use and eHealth literacy levels among university students pursuing a non–health-related degree in Pakistan. We also examined the association of the eHealth literacy levels of these young adults with their physical activity levels and dietary supplement intake.

Methods: Students from 2 leading engineering universities in Pakistan were invited to participate in a cross-sectional anonymous web-based survey in order to collect data on their internet use, eHealth literacy, and dietary supplement intake. Of the 900 eligible university students who were invited to participate, 505 (56.1%) students who completed the questionnaire were included in the analysis. The findings were converted to median values and frequency analyses were performed. The associations between the variables were determined using the chi-square test; $P \leq .05$ was considered significant.

Results: In this study, the median eHealth literacy scale (eHEALS) score was 29, which did not vary across gender. The most common type of health-related information that was searched by the participants was that related to maintaining a healthy lifestyle (305/505, 60.4%). Participants with high eHEALS scores were those who used the internet frequently for finding people with similar health issues ($P<.001$). The use of specific social media platforms was not associated with the perceived eHealth literacy levels. Neither the frequency of physical activity nor the dietary supplement use was associated with the eHealth literacy of the participants.

Conclusions: University students in non–health-related disciplines in Pakistan expressed high confidence in their skills to find health-related information on the internet, as indicated by the aggregate eHEALS scores. However, the findings of our study show that the perceived eHealth literacy was not associated with health behaviors such as physical activity and dietary supplement intake. Further research is necessary to investigate the extent to which eHealth literacy can be considered as a panacea for solving public health challenges in developing countries.

(J Med Internet Res 2020;22(6):e17014) doi:10.2196/17014

KEYWORDS

internet; digital health; eHealth; eHealth literacy; internet use; physical activity; university students; Pakistan

Introduction

The role of internet and smartphone ubiquity in health care improvement, particularly in a resource-constrained developing world, has gained global interest in the past decade [1]. Under the umbrella terms of electronic health (eHealth) and mHealth, researchers have investigated possible ways in which the proliferation of smartphone devices and internet platforms can be used to improve health education and health outcomes in the developing world [2].

In developed countries such as the United States and Japan, more than 70% of the population has reported the use of internet as the primary source of information, especially for self-care management; similarly, in developing countries, the use of internet for accessing health information has been rapidly increasing [3,4]. This trend is not surprising, considering the tech savviness of young adults and the easy accessibility to the internet. However, these web-based health interventions or programs cannot be executed as technical programs in a vacuum, ignoring the complex contexts in which they are implemented [5]. Therefore, in digital health literacy, it is important that consumer-focused electronic resources are designed such that they are aligned with the literacy levels of the intended audience. eHealth literacy is identified as an important skill set for eHealth users, and it refers to the ability of an individual to seek, find, understand, and appraise health information from electronic resources and apply such knowledge to addressing or solving a health problem [6]. The eHealth literacy scale consists of several levels, which show the set of skills that are required to effectively engage the use of information technology for improving health. The lower levels of the scale comprise operational and navigational skills, while the higher levels require the ability to choose and critically evaluate the available information [7]. eHealth literacy can help developing countries access web-based health information resources effectively. Studies examining eHealth literacy [4,8-11] have largely been performed in developed countries such as the United States, Finland, China, and Japan—with particular emphasis on young adults and college students. However, only a few studies have identified an association between eHealth literacy and health behaviors, while even fewer studies have examined the association between eHealth literacy and the health behaviors of young adults in developing countries [12].

Despite the increasing use of the internet as a source of health information in developing countries, eHealth literacy is an unexplored entity, which has been created and refined by the developed world [13]. While technology-based health information is being proposed as the solution to elevate the health literacy levels of the Pakistani population [14], there is very limited investigation of the eHealth literacy of the health seekers—with focus only on those seeking health-related education [15]. The latest National Human Development Report of Pakistan [16] highlighted that only little is known about the health literacy and the health behaviors of Pakistani youths—especially among those who are well-educated and frequent users of the internet.

Therefore, in this study, we aimed to examine the association between the patterns of internet use for seeking health information and the eHealth literacy of university students in Pakistan. Further, we investigated the possible association of eHealth literacy with the health behaviors (physical exercise and dietary supplement intake) of these university students.

Methods

Participants

Eligible participants in this study were Pakistani university students aged 18 years or above and they were enrolled in an undergraduate or a postgraduate course primarily in any degree other than in medical or health sciences. The convenience sample of the participants was approached using the university group email lists (N=900) of 2 leading engineering universities in Pakistan by one of the investigators (AB). Participants were expected to have adequate computer skills as they studied in engineering universities with sophisticated technical infrastructure.

Measurements

A quantitative cross-sectional design was used to conduct this research. Students enrolled in the universities had to meet the English language requirements prior to admission, as the mode of instruction in the anonymous web-based survey was English. Therefore, all the participants had to comprehend and answer the questions in English. The survey instrument was developed based on a review of previous studies consisting of validated items [8,17,18]. The questionnaire consisted of the following key items.

1. Demographic data and patterns of internet use: This item consisted of questions on participant age, gender, and education level. The internet use patterns were explored by asking questions on the frequency of internet use for browsing health information and common health-related topics. These questions were designed on the basis of previous studies exploring the patterns and reasons for seeking web-based health information [17]. Finally, participants rated their frequency of using Google and other social media platforms as a venue for seeking health information on a scale of 1 to 5 (1=Never, 5=Very Often). The participants rated the following social media platforms: Google, WhatsApp, Twitter, Snapchat, Instagram, and Facebook [8].
2. Perceived eHealth literacy: eHealth literacy was measured using the English version of the eHealth literacy scale (eHEALS) [18], which is an 8-item instrument measured on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). The eHEALS scores ranged from 8 to 40, wherein a high score indicates high perceived eHealth literacy and a low score indicates low perceived eHealth literacy.
3. Health behaviors: Physical activity was assessed through questions on the weekly frequency of moderate physical activity. Supplement intake was assessed through questions on the use of dietary supplements. For those who used dietary supplements, further information was requested on the type of supplement used.

The ethical approval for this study was obtained from the Human Research Ethics Committee (Approval number: 1700000897) of Queensland University of Technology. This study has been reported in accordance with the Checklist for Reporting the Results of Internet E-Surveys [19].

Survey Validation and Administration

A pilot test of the survey was completed with a sample of 14 undergraduate university students in Pakistan to ensure that the target audience understood the meaning of each question and response. The pilot feedback was used to edit the survey accordingly. The revised survey was administered using the web-based Key Survey tool. We believed that a web-based survey would be appropriate for this study because responders to such surveys can use the computer and internet effectively. The identified eligible participants were sent the link to the web-based survey via their university group email and 1 reminder was sent to give them the opportunity to complete the survey if they had not already done so. The participation in this survey was voluntary and no incentives were offered for participation. The survey remained open for 6 weeks (December 2017 to January 2018).

Data Analysis

A data matrix was produced from the completed questionnaires by using SAS 9.4 for Windows (SAS Institute Inc). Qualitative variables were expressed as numbers and percentages. The percentages were calculated based on the valid cases only. The chi-square test was performed to determine the association between the variables. $P \leq .05$ was considered significant. The total scores of the eHEALS were summed to a range from 8 to 40, with high scores representing high self-perceived eHealth literacy. We divided the eHEALS score into one of the 2 categories (high or low) relative to the median group value (median 29, IQR 26-32); we did this in accordance to that done in previous studies [4,8] that used eHEALS to analyze the associations between the frequency of seeking health information, eHealth literacy, and health behaviors.

Results

Participants’ Demographic Data and Internet Use Patterns

Of the 900 eligible students invited to participate, 559 (62.1%) students logged onto the web-based questionnaire platform and 505 (56.1%) students completed the questionnaire by providing all the demographic information; thus, 505 students were included in the final analysis (Table 1). Of these 505 students, 211 (41.8%) students were females; 88.1% (445/505) of the participants were younger than 25 years and 85.3% (431/505) of them were pursuing an undergraduate degree. We found that 79.6% (402/505) of the participants were identified as frequent users of the internet (almost every day); 11.1% (56/505) of the participants reported using the internet for searching health-related information almost every day, while 46.1% (233/505) used it once a week or more.

Table 1. Demographic data and internet use patterns of the participants (n=505).

Demographic data and frequency of internet use	Value, n (%)
Age (years)	
16-25	445 (88.1)
26-30	38 (7.5)
30+	22 (4.2)
Gender	
Male	294 (58.2)
Female	211 (41.8)
Education level	
Undergraduate degree	431 (85.3)
Postgraduate	74 (14.6)
Frequency of internet use	
Almost everyday	402 (79.6)
Several days a week (≥1 day)	88 (17.5)
Less than once a week or never	15 (3.0)
Frequency of internet use for health-related information	
Almost everyday	56 (11.1)
Several days a week (≥1 day)	233 (46.1)
Less than once a week or never	216 (42.7)

Internet Use Patterns and eHealth Literacy

With regard to eHealth literacy, only 399 participants who used the internet for finding health-related information answered the 8 standard eHEALS questions. The median eHEALS score was 29 and this value did not vary with gender or the education level. Based on the eHEALS score, the participants were divided into 2 groups: the first group had scores higher than the median eHEALS score (191/399, 47.8%) while the second group had scores lower than the median eHEALS score (208/399, 52.1%). The most common type of health-related information searched by the participants (305/505, 60.4%) was that related to healthy lifestyle (weight, exercise, etc)—this was common among both high and low eHealth literacy groups. However, those with high perceived eHealth literacy reported that they used the internet over the last 12 months more often for finding people with similar health issues (*P*<.001) (Table 2). Google was the most widely used search engine. Of the 401 participants who reported frequent use of Google, 365 (91.0%) used Google to obtain health-related information. With regard to other social media platforms, 64.0% (199/311) of the participants who used Facebook, 59.6% (171/287) of the participants who used Wikipedia, and 15.0% (40/265) of the participants who used Twitter reported that they used these platforms to obtain or share health-related information. The use of any of these specific social media platforms was not associated with the perceived eHealth literacy.

Health Behaviors and eHealth Literacy

Overall, 25.9% (119/459) of the participants reported engaging in moderate physical activity more than five times a week, while 23.5% (108/459) of the participants engaged in moderate physical activity less than once a week. The perceived level of eHealth literacy was not associated with the reported levels of physical activity (Table 3). Some form of supplement was taken by 48.7% (220/452) of the participants. Vitamin D supplements were reported as the most commonly used supplement (177/413, 42.9%). Dietary supplement use was significantly associated with gender (*P*=.002), with females more likely to use supplements as compared to males. The perceived level of eHealth literacy was not associated with the reported supplement use patterns (Table 3).

Table 2. Association between internet use patterns and eHealth literacy (n=399).

Internet use patterns and eHealth literacy	Value, n (%)		*P* value
	eHEALS[a] Low	eHEALS High	
Frequency of internet use for health-related information			.02
Several days a week or more	62 (44.3)	78 (55.7)	
Once a week or less	146 (56.4)	113 (43.6)	
Types of health-related information searched on the internet			
Healthy lifestyle	149 (50.7)	145 (49.3)	.33
Medication	69 (52.3)	63 (47.7)	.97
About a particular disease	84 (51.5)	79 (48.5)	.84
Find a care provider or hospital	26 (52.0)	24 (48.0)	.98
Peer support forums/other	12 (26.7)	29 (73.3)	.004
Frequency of using internet over the last 12 months for finding people with similar health issues			<.001
Often	43 (20.7)	76 (39.8)	
Rarely (less than once a month)	124 (59.7)	93 (48.7)	
Never	41 (19.7)	22 (11.5)	
Social media platforms used to obtain/share health information or discuss health problems			
Google			.91
No	17 (51.5)	16 (48.5)	
Yes	187 (52.5)	169 (47.5)	
Facebook			.15
No	57 (53.8)	49 (46.2)	
Yes	87 (45.1)	106 (54.9)	
Twitter			.74
No	106 (49.1)	110 (50.9)	
Yes	18 (46.2)	21 (53.8)	
Wikipedia			.54
No	57 (51.4)	54 (48.6)	
Yes	79 (47.6)	87 (52.4)	
WhatsApp			.60
No	81 (50.6)	79 (49.4)	
Yes	56 (47.5)	62 (52.5)	
Others (eg, Reddit)			.51
No	63 (50.8)	61 (49.2)	
Yes	25 (45.5)	30 (54.5)	

[a] eHEALS: eHealth literacy scale.

Table 3. Association between eHealth literacy and health behaviors.

Health behaviors	Value, n (%)		*P* value
	eHEALS[a] Low	eHEALS High	
Frequency of moderate physical activity (30-min activity)			.85
More than 5 times a week	54 (52.9)	48 (47.1)	
1-4 times a week	91 (49.7)	92 (50.3)	
Less than once a week	46 (52.3)	42 (47.7)	
Dietary supplement intake (Vitamin D and others)			.25
No	99 (54.4)	83 (45.6)	
Yes	93 (48.4)	99 (51.6)	
Types of supplements			
Vitamin D	74 (47.1)	83 (52.9)	.14
Calcium	26 (59.1)	18 (40.9)	.33
Vitamin C	25 (62.5)	15 (37.5)	.17
Vitamin B complex	12 (48.0)	13 (52.0)	.67
Iron	21 (63.6)	12 (36.4)	.17
Mineral supplements (zinc, magnesium, etc)	11 (52.4)	10 (47.6)	.98
Multivitamins (not containing vitamin D)	8 (42.1)	11 (57.9)	.37

[a] eHEALS: eHealth literacy scale.

Discussion

To our knowledge, this study is one of the very few studies that have examined the eHealth literacy levels among young adults in developing countries and the association of eHealth literacy with internet use patterns and health behaviors. After controlling for the sociodemographic variables, we found that eHealth literacy was not associated with health behaviors such as physical activity and dietary supplement intake. Although there is evidence that eHEALS may consist of several subscales, we analyzed the eHEALS as a unidimensional factor, following Hyde's recommendation [20]. Therefore, our analysis considers the full eHEALS, which also allows us to compare our results with those of other studies [8].

Our population consisted of university students pursuing an engineering degree in Pakistan; the median eHEALS score in our study was similar to that reported among young adults seeking a health-related degree in Pakistan [15]. Further, the median eHEALS score in this study was comparable to that reported previously in an adult population in the United States (mean eHEALS score, 29.0), slightly higher than that reported in the adult populations of Korea, Iran, and Kuwait (mean eHEALS scores, 28.0, 28.2, and 28.6, respectively), and much higher than that reported in an adult population in Japan (mean eHEALS score, 23.4) [4,8-10]. Similar to that reported in previous studies, most sociodemographic variables (eg, gender, age) examined in this study were not significant predictors of eHealth literacy [4,8-10]. As reported previously [4,10], it is possible that the geographical location and cultural and language barriers could affect the eHEALS scores because of limited availability of health-related information in languages other than English. This may not be the case for the population in this study as English is the official medium of instruction in the Pakistani university educational system and participants in this study are frequent users of the internet.

Unlike the findings of recent studies from the United States and Japan [4,21] that showed a significant association between eHEALS scores in young adults and health behaviors such as exercise, we found no association between eHEALS scores and the reported physical activity levels. In Pakistan, very few studies have investigated how digital health literacy can improve health behaviors. In 2018, Saeed et al [22] evaluated the digital health literacy of patients with type 2 diabetes (N=204) in Lahore, Pakistan, and they reported that patients with high socioeconomic status agreed that access to digital medical content via smartphones and tablets helped them improve their health. However, only the use of internet and the high eHealth literacy levels among educated young adults cannot translate into health behavior improvements (eg, increase in the physical activity levels) if the communities that they live in lack the needed infrastructure to facilitate the required physical activity programs. It is important to note that developing countries such as Pakistan—unlike Japan and the United States—lack culturally suitable infrastructure (eg, safe outdoor facilities for young women) and civic amenities such as playgrounds, parks, and gyms (even within universities) [23,24]. The lack of such facilities inhibits the realization of positive health behaviors and encourages indoor sedentary lifestyles [16].

We also found that there was no significant association (P=.25) between eHealth literacy and dietary supplement intake. Recent studies from developed countries have reported internet as the primary source of information to inform attitudes toward the use of dietary supplements among young adults [25,26]. Recent evidences have shown that eHealth literacy influences the food choices of adolescents [27]. In the last few years, Pakistan has witnessed a rise in the demand for nutraceuticals; dietary supplements attained a compound annual growth rate of 14% in 2016 [28]. These nutraceuticals are mostly used by the educated population [29]. In light of these interesting preliminary trends, future research should focus on the effects of the patterns of internet searches and social behaviors on supplement intake. The aspects around eHealth literacy and its association with health behaviors such as dietary supplement use have not been reported in any developing country to date, to the best of our knowledge. This study presents a valuable contribution to establish evidence, in particular, for a country facing pressures to improve health literacy and public health outcomes.

This study has several limitations. First, the participants were recruited from a convenience sample of university students in Pakistan; thus, the relationships assessed may have been biased because of the potentially non-representative nature of this sample. However, this approach was viable because it provided us with a sizable sample in a reasonable amount of time for a population wherein such investigations have not been performed previously. Our study presents the needed ground to build on future studies that explore the eHealth literacy of broad population groups so that the findings can be applied to other developing countries. Second, as in previous studies, health behavior and eHealth literacy were examined using a self-administered questionnaire. Thus, inaccuracies in estimating the health behavior and eHealth literacy level were unavoidable, and data of some of the participants were missing. In addition, we did not investigate the specific health status of the participants or whether they were searching for health-related information for another family member. Moreover, this study was limited to only examining supplement use as a dietary behavior. Future studies should explore additional dietary behaviors (eg, nutritional intake, smoking) and the context for using the internet for health-related information. Our findings revealed high eHEALS scores among the participants. This high confidence shown by the university students presents a myriad of opportunities to better engage people digitally and conveniently. As internet is being increasingly used as a source of health information, further research is needed to identify the mechanisms linking eHealth literacy with health behaviors toward designing contextually effective strategies for improving self-care in developing countries.

Acknowledgments

This research was financially supported by Queensland University of Technology Early Career and Research Development grant. We thank the administration staff in the participating universities for their assistance in questionnaire distribution and collection throughout the study period.

Authors' Contributions

AT and SK conceived and designed this study. AB and AT conducted the data collection. All authors contributed to the analysis and interpretation of data. AT wrote this manuscript with all co-authors commenting on the drafts of the paper. All authors gave their approval for the final version to be published.

Conflicts of Interest

None declared.

References

1. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. Bull World Health Organ 2012 May 1;90(5):332-340 [FREE Full text] [doi: 10.2471/BLT.11.099820] [Medline: 22589566]
2. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
3. Fox S, Duggan M. Pew Research Center - Internet & Technology. 2013. Information Triage URL: https://www.pewresearch.org/internet/2013/01/15/information-triage/ [accessed 2020-05-14]
4. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
5. Tariq A, Durrani S. One size does not fit all: The importance of contextually sensitive mHealth strategies for frontline female health workers. In: mHealth Innovation in Asia. Dordrecht (NL): Springer; 2017:7-29.
6. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9]
7. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
8. Alhuwail D, Abdulsalam Y. Assessing Electronic Health Literacy in the State of Kuwait: Survey of Internet Users From an Arab State. J Med Internet Res 2019 May 24;21(5):e11174 [FREE Full text] [doi: 10.2196/11174] [Medline: 31127723]

9. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
10. Dashti S, Peyman N, Tajfard M, Esmaeeli H. E-Health literacy of medical and health sciences university students in Mashhad, Iran in 2016: a pilot study. Electron Physician 2017 Mar;9(3):3966-3973 [FREE Full text] [doi: 10.19082/3966] [Medline: 28461871]
11. Athanasopoulou C, Välimäki M, Koutra K, Löttyniemi E, Bertsias A, Basta M, et al. Internet use, eHealth literacy and attitudes toward computer/internet among people with schizophrenia spectrum disorders: a cross-sectional study in two distant European regions. BMC Med Inform Decis Mak 2017 Sep 20;17(1):136 [FREE Full text] [doi: 10.1186/s12911-017-0531-4] [Medline: 28931385]
12. Griebel L, Enwald H, Gilstad H, Pohl AL, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2018 Dec;43(4):427-442. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
13. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center 2016;22:1-44 [FREE Full text]
14. Sabzwari SR. Health literacy in Pakistan: Exploring new ways of addressing an old challenge. J Pak Med Assoc 2017 Dec;67(12):1901-1904 [FREE Full text] [Medline: 29256538]
15. Moazzam M, Anwer F, Atique S, Kanwal S, Afzal S, Bashir A, et al. Perceived Health Literacy and Health Information Seeking Behavior among University Students in Pakistan. AMDHS 2018 Dec 30;1(1):05-07. [doi: 10.5530/amdhs.2018.1.3]
16. Shakeel A. Pakistan National Human Development Report. Unleashing the potential of a young Pakistan. Islamabad: UNDP; 2018 Jul 24. URL: http://hdr.undp.org/en/content/unleashing-potential-young-pakistan [accessed 2020-01-20]
17. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin SF, et al. Association of Online Health Information-Seeking Behavior and Self-Care Activities Among Type 2 Diabetic Patients in Saudi Arabia. J Med Internet Res 2015;17(8):e196 [FREE Full text] [doi: 10.2196/jmir.4312] [Medline: 26268425]
18. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
19. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-Factor Structure of the eHealth Literacy Scale Among Magnetic Resonance Imaging and Computed Tomography Outpatients: A Confirmatory Factor Analysis. JMIR Hum Factors 2018 Feb 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
21. Britt RK, Collins WB, Wilson K, Linnemeier G, Englebert AM. eHealth Literacy and Health Behaviors Affecting Modern College Students: A Pilot Study of Issues Identified by the American College Health Association. J Med Internet Res 2017 Dec 19;19(12):e392 [FREE Full text] [doi: 10.2196/jmir.3100] [Medline: 29258979]
22. Saeed H, Saleem Z, Naeem R, Shahzadi I, Islam M. Impact of health literacy on diabetes outcomes: a cross-sectional study from Lahore, Pakistan. Public Health 2018 Mar;156:8-14. [doi: 10.1016/j.puhe.2017.12.005] [Medline: 29353668]
23. Samir N, Mahmud S, Khuwaja AK. Prevalence of physical inactivity and barriers to physical activity among obese attendants at a community health-care center in Karachi, Pakistan. BMC Res Notes 2011 Jun 6;4(1):174. [doi: 10.1186/1756-0500-4-174]
24. Abbasi IN. Socio-cultural Barriers to Attaining Recommended Levels of Physical Activity among Females: A Review of Literature. Quest 2014 Oct 14;66(4):448-467. [doi: 10.1080/00336297.2014.955118]
25. Nishijima C, Kobayashi E, Sato Y, Chiba T. A Nationwide Survey of the Attitudes toward the Use of Dietary Supplements among Japanese High-School Students. Nutrients 2019 Jun 28;11(7):1469 [FREE Full text] [doi: 10.3390/nu11071469] [Medline: 31261616]
26. Sirico F, Miressi S, Castaldo C, Spera R, Montagnani S, Di Meglio F, et al. Habits and beliefs related to food supplements: Results of a survey among Italian students of different education fields and levels. PLoS One 2018;13(1):e0191424 [FREE Full text] [doi: 10.1371/journal.pone.0191424] [Medline: 29351568]
27. Sousa P, Frontini R, Luis L, Ferreira R. The relationship between eHealth literacy and adolescents? food choices. European Journal of Public Health 2019 Jun 06;29(Supplement_1):32-33 [FREE Full text] [doi: 10.1093/eurpub/ckz034.099]
28. Marketresearch.com.: Euromonitor International; 2016. Vitamins and dietary supplements in Pakistan URL: https://www.marketresearch.com/Euromonitor-International-v746/Vitamins-Dietary-Supplements-Pakistan-10328873/ [accessed 2020-05-14]
29. Naqvi AA, Ahmad R, Zehra F, Yousuf R, Kachela B, Nehal Nadir M. Dietary Supplement Use Among Students of Pharmacy Colleges in the City of Karachi, Pakistan: Prevalence, Opinions, and Attitudes. J Diet Suppl 2019;16(2):166-178. [doi: 10.1080/19390211.2018.1443191] [Medline: 29561199]

Abbreviations

eHEALS: eHealth literacy scale
eHealth: electronic health

Edited by G Eysenbach; submitted 11.11.19; peer-reviewed by JJ Mira, C Fernández, T Turk; comments to author 02.12.19; revised version received 22.01.20; accepted 24.01.20; published 10.06.20.

Please cite as:
Tariq A, Khan SR, Basharat A
Internet Use, eHealth Literacy, and Dietary Supplement Use Among Young Adults in Pakistan: Cross-Sectional Study
J Med Internet Res 2020;22(6):e17014
URL: http://www.jmir.org/2020/6/e17014/
doi:10.2196/17014
PMID:32519974

Original Paper

Factors Associated With the Actual Behavior and Intention of Rating Physicians on Physician Rating Websites: Cross-Sectional Study

Xi Han[1], PhD; Bei Li[2], DPH; Tingting Zhang[3], PhD; Jiabin Qu[4], PhD

[1]School of Business Administration, Guangdong University of Finance & Economics, Guangzhou, China
[2]School of Health Service Management, Southern Medical University, Guangzhou, China
[3]School of Information Engineering, Nanjing Audit University, Nanjing, China
[4]Library of Yantai University, Yantai, China

Corresponding Author:
Xi Han, PhD
School of Business Administration
Guangdong University of Finance & Economics
Luntou Road 21
Haizhu District
Guangzhou, 510320
China
Phone: 86 13512791305
Email: hanx015@163.com

Abstract

Background: Although online physician rating information is popular among Chinese health consumers, the limited number of reviews greatly hampers the effective usage of this information. To date, little has been discussed on the variables that influence online physician rating from the users' perspective.

Objective: This study aims to investigate the factors associated with the actual behavior and intention of generating online physician rating information in urban China.

Methods: A web-based cross-sectional survey was conducted, and the valid responses of 1371 Chinese health consumers were recorded. Using a pilot interview, we analyzed the effects of demographics, health variables, cognitive variables, and technology-related variables on online physician rating information generation. Binary multivariate logistic regression, multiple linear regression, one-way analysis of variance analyses, and independent samples *t* test were performed to analyze the rating behavior and the intentions of the health consumers. The survey instrument was designed based on the existing literature and the pilot interview.

Results: In this survey, 56.7% (778/1371) of the responders used online physician rating information, and 20.9% (287/1371) of the responders rated the physicians on the physician rating website at least once (posters). The actual physician rating behavior was mainly predicted by health-related factors and was significantly associated with seeking web-based physician information (odds ratio [OR] 5.548, 95% CI 3.072-10.017; *P*<.001), usage of web-based physician service (OR 2.771, 95% CI 1.979-3.879; *P*<.001), health information-seeking ability (OR 1.138, 95% CI 0.993-1.304; *P*=.04), serious disease development (OR 2.699, 95% CI 1.889-3.856; *P*<.001), good medical experience (OR 2.149, 95% CI 1.473-3.135; *P*<.001), altruism (OR 0.612, 95% CI 0.483-0.774; *P*<.001), self-efficacy (OR 1.453, 95% CI 1.182-1.787; *P*<.001), and trust in online physician rating information (OR 1.315, 95% CI 1.089-1.586; *P*=.004). Some factors influencing the intentions of the posters and nonposters rating the physicians were different, and the rating intention was mainly determined by cognitive and health-related factors. For posters, seeking web-based physician information (β=.486; *P*=.007), using web-based medical service (β=.420; *P*=.002), ability to seek health information (β=.193; *P*=.002), rating habits (β=.105; *P*=.02), altruism (β=.414; *P*<.001), self-efficacy (β=.102; *P*=.06), trust (β=.351; *P*<.001), and perceived ease of use (β=.275; *P*<.001) served as significant predictors of the rating intention. For nonposters, ability to seek health information (β=.077; *P*=.003), chronic disease development (β=.092; *P*=.06), bad medical experience (β=.047; *P*=.02), rating habits (β=.085; *P*<.001), altruism (β=.411; *P*<.001), self-efficacy (β=.171; *P*<.001), trust (β=.252; *P*<.001), and perceived usefulness of rating physicians (β=.109; *P*<.001) were significantly associated with the rating intention.

Conclusions: We showed that different factors affected the physician rating behavior and rating intention. Health-related variables influenced the physician rating behavior, while cognitive variables were critical in the rating intentions. We have proposed some practical implications for physician rating websites and physicians to promote online physician rating information generation.

(J Med Internet Res 2020;22(6):e14417) doi:10.2196/14417

KEYWORDS

online physician rating; user-generated content; physician rating website; behavioral intention; actual behavior

Introduction

With the development of user-generated content and prevalent use of mobile devices, some industries (ie, food service, travel, and e-commerce) have begun gathering web-based reviews, because of which many websites of these industries have now become reliable and effective [1]. The health care system has also started garnering web-based reviews, even though the development of the review platform was slow in the initial years. Similar to how people review products, health consumers can post their opinions on the health care received and have access to other patients' opinions on the care they received from physicians. In particular, information on physician rating websites seems to play an increasingly important role in the life of health consumers and has attracted the attention of medical practitioners. A study [2] showed that in 2007, only 3%-7% of the health consumers in the United States used physician rating websites, but the proportion increased to 23% in 2012 [3], 25% in 2013 in Germany [4], 42% in 2014 in the United States [5], and 43.6% in 2016 in Germany [6]. Online physician rating information is an important factor that seems to increasingly influence a patient's choice of medical practitioners [7]. On the contrary, physicians have always reported a negative attitude toward online physician rating information [8-10] because they fear that limited reviews could produce bias and negative web-based reviews could ruin their reputation [11,12]. Previous studies have shown the content analyses of physician rating websites in different clinical specialties [13-16], and the average number of reviews per physician was found to be low [17-19], even though the number of reviews has increased rapidly in the past few years [20]. Emmert et al [4] reported that only 11.03% of the Germans had posted ratings on a physician rating website in 2013, while this percentage increased to 23% in 2016 [6]. The limited number of reviews is the key factor that has affected the adoption of online physician rating information by both physicians and consumers. Thus, it is important to investigate the factors that predict the generation of online physician rating information from the perspective of health consumers.

Previous studies have mainly focused on the usage of online physician rating information and the related factors. Terlutter et al [21] reported that women, young adults, and people with higher education levels or chronic diseases used physician rating websites more than their counterparts. Galizzi et al [22] found that white British people and people with high incomes were less likely to use physician rating websites. Further, female participants, widows, and those with high health care utilization showed a significantly high likelihood of being aware of physician rating websites [4]. In China, due to the promotion of the "internet + health care" strategy by the government, physician rating websites are becoming increasingly popular among urban citizens. Hao et al [23-25] conducted a content analysis of Chinese physician rating websites and identified factors that were related to physician ratings. Zhang et al [26] analyzed the negative comments on physician rating websites to identify the possible solutions for improving patient satisfaction. Li et al [27] developed a hierarchical topic taxonomy to uncover the latent structure of the physician reviews and illustrated its application in mining data on patients' interests. Deng et al [28], Han et al [29], and Li and Hubner [30] investigated how web-based ratings and other factors influence the selection of physicians by the Chinese consumers. However, studies on physician rating websites in the Chinese context are still limited and little is known about the factors influencing the generation of online physician rating information.

To fill this research gap, we first conducted a web-based pilot interview and recruited 30 Chinese citizens with different education levels, occupational backgrounds, and hometowns. We introduced several Chinese physician rating websites at the beginning of the interview; thereafter, the participants reported their experience of using the physician rating websites. Only 5 of the 30 participants generated online physician rating information. Finally, participants were asked why they did or did not generate online physician rating information. Following the procedure of qualitative analysis, 3 researchers transcribed and coded the data, and we finally identified the factors related to the generation of online physician rating information. These factors were divided into three dimensions, namely, health and habit-related factors (ie, usage of web-based physician service, ability to seek health information, health conditions, experience of medical service, and rating habits in the e-commerce context), cognitive factors (ie, altruism, self-efficacy, and trust in online physician rating information), and technology-related factors (ie, perceived usefulness and perceived ease of use). Cognitive factors were often reported to be associated with knowledge-sharing behavior [31], and technology-acceptance factors were often associated with system adoption [32]. The use of physician rating websites to rate physicians signifies knowledge-sharing and system-adoption behavior. In this study, we applied a similar procedure used by Terlutter et al [21], Galizzi et al [22], and Emmert et al [4] to empirically explore the significant factors that predict the actual behavior and intention of rating the physicians on the physician rating websites. The results of this study will be useful for understanding the web-based rating behavior of health consumers and for further promoting the development of physician rating websites.

Methods

Participant Recruitment

Since physicians in rural areas are seldom rated on physician rating websites [23], our study focused on physicians in the urban regions of China. We used the snowball sampling method to recruit participants through web-based social networking. First, we selected 160 WeChat friends with varying gender, education levels, and occupational backgrounds to complete the web-based questionnaire. Second, we requested these participants to invite friends with varying genders, education levels, and occupational backgrounds to participate in the web-based survey. In total, we received 1556 responses from September 2018 to October 2018 and from August 2019 to October 2019. Among the total number of responders, 185 were excluded from the analysis because of inconsistent answer patterns (eg, flatliners, contradictions) or because the participants tried to complete the questionnaire quickly with incomplete answers in a short span of time. Finally, this study considered the responses of 1371 valid respondents. We paid each participant 2 RMB (US $0.3) to compensate for their time.

Questionnaire Design

The researchers designed a survey based on the existing literature [21,22] and their pilot interview. All items, except categorical variables, were measured using a 7-point Likert scale, with the options ranging from “strongly disagree” to “strongly agree” (Multimedia Appendix 1). To ensure the validity of the scale in our questionnaire, we adopted measurement items from the existing literature and we modified some items to adapt to the online physician rating scenario based on our first pilot study with 30 Chinese citizens. We calculated the mean values of the multiple items as predictor scores after checking the measurement’s internal reliability.

The questionnaire was created in English. One researcher translated it into Chinese, and then another researcher translated it back into English to ensure the consistency of the content. After developing the Chinese questionnaire, 3 researchers in health informatics were invited to assess the ease of understanding, logical consistencies, item sequence, and contextual relevance of the items in the questionnaire. We made some minor modifications based on their suggestions. Furthermore, a pilot test was conducted with 20 participants, and the items were modified slightly.

Measurements

Rating Behavior, Rating Intention, and Demographic Variables

To ensure that the respondents understood the online physician rating system, a screenshot of a physician rating website was presented in the introductory phase before the respondents answered the questions. The actual behavior of rating a physician was assessed by asking if the respondents had rated physicians on physician rating websites previously (0=no, 1=yes). We defined participants as posters if they had rated a physician on a physician rating website at least once and we defined participants as nonposters if they had never rated a physician on the physician rating website. The intention of rating a physician was assessed using a 3-item scale adapted from the study of Ajzen [33]. This scale was found to be reliable (mean 5.064, SD 1.189; Cronbach α=.949). Additionally, data on demographic variables such as age, gender, education level, marital status, monthly income, daily internet use, and the number of vulnerable family members were also collected (Multimedia Appendix 1).

Health and Habit-Related Variables

The usage of web-based physician information was assessed by asking respondents if they had ever sought physician information on the internet (0=no, 1=yes). The usage of web-based physician service was assessed by asking participants if they had ever booked or consulted a physician on the internet (0=no, 1=yes). The ability to seek health information was assessed using 2 items adapted from the model developed by Richard et al [34]. The scale (mean 4.649, SD 1.350; Cronbach α=.890) assessed the participant’s ability to search for web-based health information. The health conditions of the participants or their family members were measured using the following 2 questions: “Did you or your family members develop any chronic disease in the past 2 years?” and “Did you or your family members develop any serious disease in the past 2 years?” (0=no, 1=yes). Medical experience was measured using the following questions: “I had a very good medical experience in the past 2 years” and “I had a very bad medical experience in the past 2 years” (0=no, 1=yes). In the e-commerce era, some consumers have the habit of posting reviews after performing web-based transactions, and the habit was found to be effective in the online physician rating scenario in our pilot interview. The following rating habit in the e-commerce context was assessed using an item adapted from a previous study [35]: “Rating the product after a web-based transaction has become a habit for me” (1=strongly disagree, 7=strongly agree).

Cognitive Variables

Altruism is a behavior intended to benefit another, even when this action may involve sacrificing one’s welfare [36]. A 3-item scale on altruism was adapted from previous studies [37,38] and applied in our pilot interview. This scale was found to be reliable (mean 5.438, SD 1.042; Cronbach α=.910). Self-efficacy refers to the belief or the estimate of an individual about his/her own ability to perform a particular task [39]. The self-efficacy scale was adapted from prior studies and applied in our pilot interview [40,41], and the scale was found to be reliable (mean 4.594, SD 1.202; Cronbach α=.770). Trust refers to a situation in which one party willingly relies on the actions of the other party [42]. The scale for trust in online physician rating information was adapted from a previous study [43], and this scale was also found to be reliable (mean 4.711, SD 1.224; Cronbach α=.885).

Technology-Related Variables

Perceived usefulness refers to the usefulness of using physician rating websites to rate physicians, and perceived ease of use refers to the ease of using physician rating websites to rate physicians. Based on our pilot interview and previous studies [44,45], we adapted reliable scales for perceived usefulness (mean 5.147, SD 1.162; Cronbach α=.842) and perceived ease of use (mean 4.405, SD 1.127; Cronbach α=.742).

Data Analysis

Data were downloaded from the web-based questionnaire database to computers in our laboratory in Nanjing University, China. Two independent research assistants examined the data and removed 185 unqualified cases. Data analyses were conducted using the SPSS 23.0 software (IBM Corp). We examined the descriptive statistics for all the variables. Since we focused on participants who were aware of physician rating websites before completing our survey, binary logistic regression analysis was performed to examine the effect of the variables on the likelihood of generating online physician rating information. Multiple linear regression, one-way analysis of variance (ANOVA), and independent samples *t* test (two-tailed) were performed to investigate the different factors influencing the physician rating intentions of the posters and nonposters. We applied data screening procedures to identify problematic patterns within the data set before performing linear regression. Linear relationship, multivariate normality, multicollinearity tested by variance inflation factors (VIFs), autocorrelations tested by Durbin-Watson, and homoscedasticity were tested, and we found that the data could be used for further linear regression. A bootstrapping procedure (with 5000 bootstrap samples) was used in our regression models.

Results

Demographic Data of the Participants

The demographic characteristics of the participants are presented in Table 1. The age range of most of the participants was between 25 and 40 years. Out of the 1371 participants, 789 (57.6%) were women and 980 (71.5%) were married. The monthly income of 69.1% (947/1371) of the participants ranged between 3000 RMB (US $435) and 12,000 RMB (US $1740). With respect to the education level, 68.5% (939/1371) of the participants had completed college or higher level of education. Of the 1371 participants, 778 (56.7%) used online physician rating information and 287 (20.9%) rated the physicians on physician rating websites.

Table 1. Demographic characteristics of the sample (n=1371).

Demographic characteristics	Value, n (%）
Age (years)	
≤24	160 (11.7)
25-30	251 (18.3)
31-35	278 (20.3)
36-40	242 (17.7)
41-45	204 (14.9)
≥46	236 (17.2)
Gender	
Female	789 (57.5)
Male	582 (42.5)
Income (RMB ¥)[a]	
≤3000	164 (12.0)
3001-6000	337 (24.6)
6001-9000	326 (23.8)
9001-12,000	284 (20.7)
≥12,001	260 (19.0)
City level	
County/bureau level	337 (24.6)
Provincial level	402 (29.3)
Metropolitan	632 (46.1)
Education	
Middle school	432 (31.5)
Undergraduate	527 (38.4)
Postgraduate	412 (30.1)
Marital status	
Unmarried	391 (28.5)
Married	980 (71.5)
Children and elders	
0	240 (17.5)
1	235 (17.1)
2	317 (23.1)
3	209 (15.2)
4	178 (13.0)
≥5	192 (14.0)
Daily internet use	
T[b]≤3 h	158 (11.5)
3<T≤5 h	303 (22.1)
5<T≤7 h	292 (21.3)
7<T≤9 h	294 (21.4)
9<T≤11 h	324 (23.6)
Online physician rating awareness	

Demographic characteristics	Value, n (%）
Aware	972 (70.9)
Unaware	399 (29.1)
Online physician rating usage	
Nonusers	593 (43.3)
Users	778 (56.7)
Online physician rating generation	
Nonposters	1084 (79.1)
Posters	287 (20.9)

[a] A currency exchange rate of RMB ¥1=US $0.14 is applicable.

[b] T: Time.

Factors Associated With the Actual Behavior of Rating Physicians on Physician Rating Websites

We focused on participants who were aware of physician rating websites before our survey (n=972), and Table 2 and Table 3 show the results of 4 binary logistic regressions. In the first step, a binary logistic regression between having rated a physician or not (yes=1, no=0) as the criterion and demographic variables was performed (Model 1). Model 1 was significant. Age (β=.146; *P*=.06), monthly income (β=.197; *P*=.003), and education level (β=–.308; *P*=.02) were significantly associated with the likelihood of rating a physician on physician rating websites. However, we also noticed that Nagelkerke R^2 (.041) was quite low, and –2 log-likelihood (1130.138) was high. The results indicated that demographic variables only explained a small part of the actual rating behavior.

Then, we entered the health- and habit-related variables into regression Model 2. The model improved with a Nagelkerke R^2 change of 0.238. The regression coefficients were significant for the following variables: experience of seeking physician information on the internet (β=1.713; *P*<.001), usage of web-based physician service (β=1.019; *P*<.001), ability to seek health information (β=.129; *P*=.04), development of serious diseases (β=.993; *P*<.001), and good medical experience (β=.765; *P*<.001). We also noticed that gender (β=.410; *P*=.02) and marital status (β=–.441; *P*=.047) were significant factors associated with the actual rating behavior, while age (β=.116; *P*=.16) was not significant after health-related factors were considered.

Following Model 2, cognitive variables were entered into Model 3, which were also significant (*P*<.001). Furthermore, Model 3 showed a minor improvement over Model 2, with increased Nagelkerke R^2 change of 0.035. The significant factors in Model 2 mentioned above were still significant. Altruism was negatively (β=–.492; *P*<.001) related to the likelihood of rating physicians. Self-efficacy (β=.374; *P*<.001) and trust in online physician rating information (β=.274; *P*=.004) were significantly and positively related to the likelihood of the rating behavior.

Based on Model 3, technology-related variables were entered into Model 4. However, no improvement was observed in Nagelkerke R^2 and the regression coefficients of perceived usefulness (*P*=.42) and perceived ease of use (*P*=.33) were not significant. Significant variables in Model 2 and Model 3 were also significant in Model 4.

Besides the reliability indices mentioned above, collinearity statistics using VIF and tolerance values were tested. The results showed that VIF scores did not exceed 2.479, and tolerance values were not lower than 0.411. According to the criteria proposed by Montgomery and Peck [46] that VIF should be lower than 10 and tolerance value should be more than 0.1, our results indicated that multicollinearity was not a big concern.

Table 2. Binary logistic regressions for online physician rating behavior (Model 1-2).

	Model 1[a]				Model 2[b]			
	β	Sig.[c]	OR[d]	95% CI of OR	β	Sig.	OR	95% CI of OR
Constant	–1.267	.002	0.293	—[e]	–3.612	<.001	0.027	—
Demographic variables								
Age	.146	.06	1.138	0.990-1.308	.116	.16	1.123	0.954-1.322
Gender	.190	.21	1.209	0.896-1.633	.410	.01	1.507	1.076-2.110
Monthly income	.197	.003	1.218	1.071-1.384	.183	.01	1.201	1.041-1.387
Marital status	–.257	.21	0.773	0.518-1.155	–.441	.047	0.644	0.411-1.009
Number of children and elders	.080	.09	1.084	0.985-1.192	.042	.44	1.043	0.938-1.159
Education level	–.308	.02	0.735	0.564-0.958	–.700	<.001	0.496	0.365-0.674
Daily internet use	–.046	.40	0.955	0.858-1.063	–.086	.15	0.917	0.815-1.033
Health- and habit-related variables								
Seeking physician information	—	—	—	—	1.713	<.001	5.548	3.072-10.017
Usage of web-based medical service	—	—	—	—	1.019	<.001	2.771	1.979-3.879
Health information seeking ability	—	—	—	—	.129	.04	1.138	0.993-1.304
Chronic disease	—	—	—	—	–.104	.54	0.901	0.646-1.258
Serious disease	—	—	—	—	.993	<.001	2.699	1.889-3.856
Good medical experience	—	—	—	—	.765	<.001	2.149	1.473-3.135
Bad medical experience	—	—	—	—	.267	.13	1.306	0.926-1.843
Rating habit	—	—	—	—	.028	.59	1.029	0.928-1.140

[a] χ^2 /Sig.: 27.887 (*df*=7) /<.001; –2 log-likelihood: 1130.138; Nagelkerke R^2: 0.041.

[b] χ^2 /Sig.: 210.132 (*df*=15) /<.001; –2 log-likelihood: 947.892; Nagelkerke R^2: 0.279.

[c] Sig.: significance probability.

[d] OR: odds ratio.

[e] Not available.

Table 3. Binary logistic regressions for online physician rating behavior (Model 3-4).

	Model 3[a]				Model 4[b]			
	β	Sig.[c]	OR[d]	95% CI of OR	β	Sig.	OR	95% CI of OR
Constant	–3.273	<.001	0.038	—[e]	–3.405	<.001	0.033	—
Demographic variables								
Age	.141	.09	1.151	0.976-1.359	.139	.10	1.149	0.972-1.358
Gender	.395	.02	1.484	1.053-2.092	.408	.02	1.504	1.065-2.123
Monthly income	.158	.04	1.171	1.009-1.358	.161	.03	1.174	1.011-1.364
Marital status	–.526	.02	0.591	0.373-0.937	–.527	.03	0.590	0.371-0.940
Number of children and elders	.019	.73	1.019	0.915-1.136	.019	.74	1.019	0.914-1.135
Education level	–.763	<.001	0.466	0.338-0.643	–.764	<.001	0.466	0.337-0.645
Daily internet use	–.068	.27	0.935	0.829-1.054	–.068	.27	0.934	0.828-1.054
Health and habit-related variables								
Seeking physician information	1.810	<.001	6.113	3.355-11.137	1.812	<.001	6.121	3.357-11.159
Usage of web-based medical service	.951	<.001	2.589	1.838-3.647	.939	<.001	2.559	1.814-3.610
Health information seeking ability	.147	.04	1.158	1.009-1.330	.150	.04	1.162	1.010-1.336
Chronic disease	–.111	.52	0.895	0.638-1.256	–.117	.499	0.889	0.633-1.249
Serious disease	.958	<.001	2.607	1.812-3.751	.979	<.001	2.662	1.845-3.842
Good medical experience	.800	<.001	2.227	1.515-3.273	.788	<.001	2.200	1.372-2.849
Bad medical experience	.325	.07	1.384	0.971-1.971	.322	.07	1.380	0.968-1.967
Rating habit	–.070	.23	0.933	0.831-1.046	–.077	.20	0.926	0.823-1.042
Cognitive variables								
Altruism	–.492	<.001	0.612	0.483-0.774	–.437	.002	0.646	0.489-0.853
Self-efficacy	.374	<.001	1.453	1.182-1.787	.383	.001	1.466	1.167-1.843
Trust	.274	.004	1.315	1.089-1.586	.263	.007	1.301	1.073-1.577
Technology-related variables								
Perceived usefulness	—	—	—	—	–.102	.42	0.903	0.703-1.160
Perceived ease of use	—	—	—	—	.087	.33	1.091	0.917-1.298

[a] χ^2 /Sig.: 240.174 (*df*=18) / <.001; –2 log-likelihood: 917.851; Nagelkerke R^2: 0.314.

[b] χ^2 /Sig.: 241.789 (*df*=20) / <.001; –2 log-likelihood: 916.236; Nagelkerke R^2: 0.316.

[c] Sig.: significance probability.

[d] OR: odds ratio.

[e] Not available.

Predictive Factors for the Intention of Rating a Physician on Physician Rating Websites

To investigate how the variables influence the rating intention of the participants differently, we divided the samples into 2 groups, namely, posters group and nonposters group. Using hierarchical multiple regression analyses, we tested the effects of different dimensions of the factors on the rating intention. By controlling the demographic variables, we found that health-, cognitive-, and technology-related variables explained 21.3%, 38.1%, and 5.5% of the increased variance in the rating intention of the posters and 12.8%, 48.1%, and 0.4% of the increased variance for nonposters, respectively. The VIF and tolerance values showed that multicollinearity was not a concern in any model.

Table 4 displays the final models. For posters who rated the physicians on the physician rating websites, health and habit-related variables, that is, seeking physician information on the internet (β=.486; *P*=.007), using web-based medical services (β=.420; *P*=.002), ability to seek health information (β=.193; *P*=.002), and habits of ratings (β=.105; *P*=.02) were found to be significantly and positively related to the rating intention. The cognitive variables, that is, altruism (β=.414; *P*<.001), self-efficacy (β=.102; *P*=.06), and trust in online physician rating information (β=.351; *P*<.001) were also significant predictors of the rating intention. Perceived usefulness was not significantly associated with the rating intention (β=–.031; P=.63), while perceived ease of use (β=.271; P<.001) was a significant predictor.

For nonposters who did not rate the physicians on the physician rating websites, usage of web-based medical service (β=.077; P=.003), development of chronic disease (β=.092; P=.06), bad medical experience (β=.047; P=.02), and habits of ratings (β=.085; P<.001) were found to be significantly associated with the rating intention. Similar to that observed in the posters group, altruism (β=.411; P<.001), self-efficacy (β=.171; P<.001), and trust (β=.252; P<.001) were also found to be the predictors of the rating intentions of nonposters. Since nonposters did not post web-based physician reviews, perceived ease of use (β=.017; P=.505) was not significantly associated with the rating intention, but perceived usefulness (β=.109; P=.001) was a significant predictor of the rating intention.

Table 4. Linear regressions of the rating intentions of posters and nonposters.

	Posters[a] (n=287)				Nonposters[b] (n=1084)			
			95% CI				95% CI	
	β	Sig.[c]	Lower	Upper	β	Sig.	Lower	Upper
Constant	–.636	.045	–1.258	–0.014	.129	.49	–0.238	0.496
Demographic variables								
Age	–.091	.03	–0.172	–0.010	.015	.51	–0.030	0.061
Gender	.148	.08	–0.018	0.313	–.004	.93	–0.099	0.090
Monthly income	.018	.64	–0.056	0.091	–.005	.80	–0.045	0.034
Marital status	.004	.90	–0.052	0.059	.016	.37	–0.019	0.050
Number of children and elders	.180	.13	–0.054	0.414	–.063	.31	–0.184	0.058
Education level	–.016	.56	–0.070	0.038	–.004	.81	–0.034	0.026
Daily internet use	.010	.89	–0.142	0.162	–.004	.93	–0.091	0.083
Health- and habit-related variables								
Seeking physician information	.486	.007	0.134	0.838	–.075	.14	–0.173	0.024
Usage of web-based medical service	.420	.002	0.157	0.684	.030	.53	–0.065	0.126
Health information seeking ability	.193	.002	0.071	0.316	.077	.003	0.025	0.128
Chronic disease	–.013	.87	–0.168	0.142	.092	.06	–0.004	0.188
Serious disease	.036	.65	–0.122	0.194	.015	.86	–0.156	0.187
Good medical experience	–.005	.96	–0.208	0.198	.000	.99	–0.094	0.093
Bad medical experience	.034	.68	–0.124	0.191	.047	.02	0.021	0.232
Rating habit	.105	.02	0.018	0.193	.085	<.001	0.055	0.114
Cognitive variables								
Altruism	.414	<.001	0.294	0.535	.411	<.001	0.349	0.474
Self-efficacy	.102	.061	–0.005	0.208	.171	<.001	0.116	0.225
Trust	.351	<.001	0.260	0.441	.252	<.001	0.202	0.301
Technology-related variables								
Perceived usefulness	–.031	.635	–0.159	0.097	.109	.001	0.047	0.171
Perceived ease of use	.275	<.001	0.199	0.350	.017	.505	–0.032	0.066

[a] F /Sig.: 123.812 (df=20)/ <.001; Nagelkerke R^2: 0.617.

[b] F /Sig.: 114.296 (df=20)/ <.001; Nagelkerke R^2: 0.623.

[c] Sig.: significance probability.

Furthermore, we used one-way ANOVA and independent samples t test to compare the differences between posters (n=287) and nonposters (n=1084) on the rating intention and related factors. Following the suggestion by Fritz et al [47], Cohen d was used to estimate the effect size. It can be seen from Table 5 that posters had a higher level of rating intention than nonposters (t_{1369}=5.569; P<.001). Regarding self-efficacy, the 2 groups differed as expected (t_{1369}=5.771; P<.001), with posters ascribing higher self-efficacy than the nonposters. Table 5 also indicated that posters trusted the information on physician rating websites to a greater extent than nonposters (t_{1369}=5.549;

$P<.001$). The 2 groups did not differ significantly in altruism at $P<.05$ level ($t_{1369}=1.697$; $P=.09$). Additionally, posters perceived higher levels of usefulness ($t_{1369}=3.020$; $P=.003$) and ease of use ($t_{1369}=3.928$; $P<.001$) than nonposters. With regard to the effect size, a Cohen *d* value of 0.2 indicated a small effect and a value of 0.5 indicated a medium effect. Thus, the effect sizes for rating intention, self-efficacy, and trust were found to be medium, while the effect sizes for perceived usefulness and perceived ease of use were found to be small.

Table 5. Differences between posters and nonposters in the rating intention and related factors.

Variables	Poster, mean (SD)	Nonposter, mean (SD)	*t (df)*	*P* value	Cohen *d*
Rating intention	5.495 (1.120)	5.064 (1.189)	5.569 (1369)	<.001	0.373
Altruism	5.556 (1.074)	5.438 (1.042)	1.697 (1369)	.09	0.112
Self-efficacy	5.052 (1.229)	4.594 (1.202)	5.771 (1369)	<.001	0.377
Trust	5.150 (1.058)	4.711 (1.224)	5.549 (1369)	<.001	0.358
Perceived usefulness	5.380 (1.150)	5.147 (1.162)	3.020 (1369)	.003	0.202
Perceived ease of use	4.700 (1.085)	4.405 (1.127)	3.928 (1369)	<.001	0.267

Discussion

Principal Findings

Previous studies on online physician rating information mainly focused on the usage of online physician ratings and related factors [2-4,21], and only 2 studies [4,6] have shown the proportion (11% and 23%) of people who rated the physicians on physician rating websites. Our study focused on the urban Chinese population and found that 20.9% (287/1371) of the respondents rated the physicians on physician rating websites at least once. An important aspect of our study was that we investigated the factors that predicted the actual behavior of rating the physicians on the physician rating websites. Since only 56.7% (778/1371) of the participants had used online physician rating information, we examined the effects of different factors on the rating intentions of the posters and nonposters. Our results also show that the factors affecting the actual rating behavior and rating intention were different, even though the rating behavior was positively related to the rating intention in our partial correlation analysis ($r=.148$; $P<.001$).

Our study shows that only sociodemographic variables cannot produce a satisfying model to predict the actual behavior of rating physicians on the physician rating websites. Even though monthly income and education level were significantly correlated with the rating behavior (Model 1 in Table 2), the Nagelkerke R^2 (0.041) of the logistic regression model was low. We also found that gender and marital status were significantly associated with the rating behavior when health and cognitive variables were included. The change in Nagelkerke R^2 indicated that it was necessary to integrate additional health and cognitive variables to predict the rating behavior to a more satisfying extent. Health-related factors played an important role in predicting the likelihood of the rating behavior. In our study, participants with the experience of seeking physician information on the internet, who used web-based physician service, and who had higher ability to seek health information were more prone to rate physicians on the physician rating websites. Since there have been incidents of poor physician-patient relationships and severe cases of vicious attacks on medical professionals particularly in China [48], many health consumers choose to check physician information on the internet and seek web-based health information to avoid unpleasant medical experiences. Seeking web-based health information increased their awareness of the online physician rating information and motivated them to rate physicians. Development of serious diseases and good medical experience were also predictors of the rating behavior. This result corroborated that of previous studies that showed a large number of positive reviews on physician rating websites [49-52]. Further, altruism was negatively related to the rating behavior, indicating that egoistic motivation played a role, and nonposters showed exaggerated level of altruism in their behavior of generating online physician rating information. Self-efficacy reflects an optimistic self-belief that one can perform a task, and it was found to be positively related to the rating behavior. In a web-based context, trust is always a big concern, and it was found to be positively related to the usage of online physician rating information [21]. In our study, trust in online physician rating information was also positively related to the rating behavior. However, as most participants had not used physician rating websites, perceived usefulness and perceived ease of use were not significantly associated with the rating behavior.

Regarding the rating intention, cognitive factors explained the largest variance in the rating intention, and factors influencing the rating intention of posters and nonposters were different. The common factors were health information-seeking ability, rating habit, altruism, self-efficacy, and trust in online physician rating information. However, most health- and technology-related variables that predicted the rating intentions of the posters and nonposters were different. For health-related variables, the rating intention of the posters was mainly predicted by the usage of web-based health information or service, while the rating intention of the nonposters was associated with the health status and medical experience. Although the results in Table 2 indicated that serious disease development and good medical experience predicted the actual rating behavior, our linear regression model demonstrated that chronic disease and bad medical experience were associated with the rating intentions of the nonposters after they became aware of physician rating websites. Additionally, perceived usefulness was associated with the rating intention of the nonposters and perceived ease of use was associated with the rating intention of the posters. Further, we noticed that the posters judged the rating intention, altruism, trust in online physician rating information, perceived usefulness, and perceived ease of use higher than the nonposters.

Practical Implications

Based on the results in our study, we have some recommendations for physician rating websites and physicians who are the stakeholders of online physician rating information generation. Commercial physician rating websites are the main sources for health consumers to access online physician rating information; thus, a large amount of online physician rating information is necessary and critical for the development of physician rating websites. Large amount of online physician rating information can be generated as follows. First, physician rating websites need to cooperate with widely used search engines and social media to increase the awareness of these websites among health consumers, since our results indicated that many consumers were unaware of these websites, and usage of web-based physician information could improve online physician rating information generation behavior and rating intention. Although Chinese physician rating websites have provided services for many years and are top-ranked in the search engine results page, most health consumers are still unfamiliar with these physician rating websites and are uncertain about their reliability. Second, physician rating websites need to cooperate with hospitals officially. Health consumers have high levels of trust in public hospitals. Thus, cooperation with hospitals would enhance consumers' trust and improve the usability of commercial physician rating websites. Our findings suggested that trust was positively related to the physician rating behavior and intention. In fact, reviews on some physician rating websites in China increased greatly after physician rating websites provided booking services for hospitals. These physician rating websites are becoming increasingly popular among health consumers in cities. Third, physician rating websites must provide additional incentives for health consumers to generate online physician rating information. Knowledge sharing is an altruistic behavior. However, our results indicated altruism to be negatively related to the actual rating behavior. Egoism may play an important role in the actual rating behavior. Thus, a better incentive mechanism is needed for attracting health consumers to rate physicians on the physician rating websites. Fourth, physician rating websites need to cooperate with physicians and provide web-based medical services, besides online physician rating information. In the past 2 years, many physicians in China have begun to use physician rating websites to provide web-based medical services, which have greatly increased the number of reviews and the usability of these websites.

The results of our study could also be interesting for physicians. Online physician rating information is important for physicians to boost their reputation and to enjoy success in their careers. Thus, physicians need to actively encourage patients to generate online physician ratings by performing the following measures. First, physicians should be concerned about patients' medical experiences. We found that good medical experience predicted the actual behavior of rating the physician on the physician rating websites. This finding is consistent with that reported in previous studies that showed positive reviews for physicians [20,48]. Physicians need not worry about negative reviews ruining their reputation, even though a bad medical experience was positively related to the rating intention of nonposters. Physicians are encouraged to show empathy to their patients, who may consequently provide positive reviews about them. Second, physicians should recommend physician rating websites to their patients and encourage them to provide online physician ratings after receiving the medical service. Physician recommendations would increase patients' trust in online physician rating information and directly lead to the generation of more reviews. Even though it is embarrassing to be rated by patients, physicians should accept that online physician rating information could lead to their medical service improvement.

Limitations and Future Direction

This study has some limitations. First, we used a snowball sampling method and focused on well-educated people who were younger than 46 years in China. There is a possibility of selection bias among respondents, even though they are the potential online physician rating information users. Thus, a large randomized sample would certainly be desirable in future studies. Second, we only tested the altruistic motivation, which was found to be negatively related to the rating behavior. Future studies should analyze how egoistic motivation directly affects the rating behavior. Third, we did not differentiate people with bad medical experience from people with good medical experience. Medical experience could be an interesting variable to focus on, considering the special patient-physician relationship in China. Researchers should explore if the kind of medical experience has nuanced the effect on the intention to post online physician rating information with regard to the unsatisfying physician-patient relationships in China. Finally, factors influencing the actual behavior and intention of rating physicians were quite different in our study. Since many participants were unaware of physician rating websites before our survey, it would be better to examine how these factors affect their actual rating behavior. Even though the intention is predictive of future behavior, the self-reported intention might be exaggerated. A long-term follow-up study is needed in the future to investigate how different factors affect the actual rating behavior after health consumers become aware of the online physician rating information.

Conclusion

Since the limited number of web-based reviews greatly hampers the effective usage of physician rating information, it is important to discuss the variables that influence the generation of physician rating information from the health consumer's perspective. Our cross-sectional study shows that factors affecting the physician rating behavior and rating intention are different. We found that health-related variables influenced the physician rating behavior while cognitive variables were critical in the rating intentions. Based on our findings, we have provided some practical suggestions for physician rating websites and physicians to promote the generation of online physician rating information.

Acknowledgments

The research was supported by the Philosophy and Social Science Foundation of the Guangdong Province (GD19YTS01), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (20YJCZH039), Guangdong Basic and Applied Basic Research Joint Foundation (2019A1515110347), and Guangzhou Philosophy and Social Sciences Foundation (2020GZQN38). We sincerely appreciate the research participants in this study. We also thank the editors and the anonymous reviewers for their thoughtful comments and constructive suggestions during the review process.

Authors' Contributions

XH was the project leader and a major contributor to drafting the manuscript. BL was responsible for data collection and was another major contributor to drafting the manuscript. TZ and JQ were responsible for data collection and data analysis. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.
[DOCX File , 360 KB - jmir_v22i6e14417_app1.docx]

References

1. Pasternak A, Scherger J. Online reviews of physicians: what are your patients posting about you? Fam Pract Manag 2009;16(3):9-11 [FREE Full text] [Medline: 19492765]
2. Tu H, Lauer J. Word of mouth and physician referrals still drive health care provider choice. Res Brief 2008 Dec(9):1-8. [Medline: 19054900]
3. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
4. Emmert M, Meier F, Pisch F, Sander U. Physician Choice Making and Characteristics Associated With Using Physician-Rating Websites: Cross-Sectional Study. J Med Internet Res 2013 Aug 28;15(8):e187. [doi: 10.2196/jmir.2702]
5. Leslie J. Software Advice. 2014 Nov 19. Patient Use of Online Reviews URL: http://www.softwareadvice.com/medical/industryview/online-reviews-report-2014/ [accessed 2019-04-23]
6. McLennan S, Strech D, Meyer A, Kahrass H. Public Awareness and Use of German Physician Ratings Websites: Cross-Sectional Survey of Four North German Cities. J Med Internet Res 2017 Nov 09;19(11):e387. [doi: 10.2196/jmir.7581]
7. McBride DL. Parental use of online physician rating sites. J Pediatr Nurs 2015;30(1):268-269. [doi: 10.1016/j.pedn.2014.10.003] [Medline: 25450443]
8. Wangler J, Jansky M. How Family Practitioners Judge Physician Rating Websites - Assessments, Experiences, Effects. ZFA 2017;93(5):216-220. [doi: 10.3238/zfa.2017.0216-0220]
9. Samora JB, Lifchez SD, Blazar PE, American Society for Surgery of the Hand EthicsProfessionalism Committee. Physician-Rating Web Sites: Ethical Implications. J Hand Surg Am 2016 Jan;41(1):104-110.e1. [doi: 10.1016/j.jhsa.2015.05.034] [Medline: 26304734]
10. Holliday AM, Kachalia A, Meyer GS, Sequist TD. Physician and Patient Views on Public Physician Rating Websites: A Cross-Sectional Study. J GEN INTERN MED 2017 Feb 1;32(6):626-631. [doi: 10.1007/s11606-017-3982-5] [Medline: 28150098]
11. Gordon HS. Consider Embracing the Reviews from Physician Rating Websites. J Gen Intern Med 2017 Jun;32(6):599-600 [FREE Full text] [doi: 10.1007/s11606-017-4036-8] [Medline: 28299603]
12. Ellimoottil C, Leichtle S, Wright C, Fakhro A, Arrington A, Chirichella T, et al. Online physician reviews: the good, the bad and the ugly. Bull Am Coll Surg 2013 Sep;98(9):34-39. [Medline: 24455818]
13. McLennan S. Quantitative Ratings and Narrative Comments on Swiss Physician Rating Websites: Frequency Analysis. J Med Internet Res 2019 Jul 26;21(7):e13816 [FREE Full text] [doi: 10.2196/13816] [Medline: 31350838]
14. Liu J, Hou S, Evans R, Xia C, Xia W, Ma J. What Do Patients Complain About Online: A Systematic Review and Taxonomy Framework Based on Patient Centeredness. J Med Internet Res 2019 Aug 07;21(8):e14634 [FREE Full text] [doi: 10.2196/14634] [Medline: 31392961]
15. Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL. What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews. J Med Internet Res 2019 Apr 8;21(4):e12521. [doi: 10.2196/12521]
16. Pike CW, Zillioux J, Rapp D. Online Ratings of Urologists: Comprehensive Analysis. J Med Internet Res 2019 Jul 02;21(7):e12436 [FREE Full text] [doi: 10.2196/12436] [Medline: 31267982]

17. Kirkpatrick W, Abboudi J, Kim N, Medina J, Maltenfort M, Seigerman D, et al. An Assessment of Online Reviews of Hand Surgeons. Arch Bone Jt Surg 2017 May;5(3):139-144 [FREE Full text] [Medline: 28656160]
18. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association Between Physician Online Rating and Quality of Care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
19. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website Characteristics and Physician Reviews on Commercial Physician-Rating Websites. JAMA 2017 Feb 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]
20. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015 Jan 12;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
21. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
22. Galizzi MM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ Open 2012;2(6):e001493 [FREE Full text] [doi: 10.1136/bmjopen-2012-001493] [Medline: 23148340]
23. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
24. Hao H, Zhang K. The Voice of Chinese Health Consumers: A Text Mining Approach to Web-Based Physician Reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
25. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: International comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007]
26. Zhang W, Deng Z, Hong Z, Evans R, Ma J, Zhang H. Unhappy Patients Are Not Alike: Content Analysis of the Negative Comments from China's Good Doctor Website. J Med Internet Res 2018 Jan 25;20(1):e35 [FREE Full text] [doi: 10.2196/jmir.8223] [Medline: 29371176]
27. Li J, Liu M, Li X, Liu X, Liu J. Developing Embedded Taxonomy and Mining Patients' Interests From Web-Based Physician Reviews: Mixed-Methods Approach. J Med Internet Res 2018 Aug 16;20(8):e254 [FREE Full text] [doi: 10.2196/jmir.8868] [Medline: 30115610]
28. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The Effect of Online Effort and Reputation of Physicians on Patients' Choice: 3-Wave Data Analysis of China's Good Doctor Website. J Med Internet Res 2019 Mar 08;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
29. Han X, Qu J, Zhang T. Exploring the impact of review valence, disease risk, and trust on patient choice based on online physician reviews. Telemat Informatics 2019 Dec;45:101276. [doi: 10.1016/j.tele.2019.101276]
30. Li S, Hubner A. The Impact of Web-Based Ratings on Patient Choice of a Primary Care Physician Versus a Specialist: Randomized Controlled Experiment. J Med Internet Res 2019 Jun 28;21(6):e11188 [FREE Full text] [doi: 10.2196/11188] [Medline: 31254337]
31. Ma WW, Chan A. Knowledge sharing and social media: Altruism, perceived online attachment motivation, and perceived online relationship commitment. Comput Human Behav 2014 Oct;39:51-58. [doi: 10.1016/j.chb.2014.06.015]
32. Szajna B. Software Evaluation and Choice: Predictive Validation of the Technology Acceptance Instrument. MIS Q 1994 Sep;18(3):319-324. [doi: 10.2307/249621]
33. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
34. Richard M, Chebat J, Yang Z, Putrevu S. A proposed model of online consumer behavior: Assessing the role of gender. J Bus Res 2010 Sep;63(9-10):926-934. [doi: 10.1016/j.jbusres.2009.02.027]
35. Shiau W, Luo MM. Continuance intention of blog users: the impact of perceived enjoyment, habit, user involvement and blogging time. Behav Inf Technol 2013 Jun;32(6):570-583. [doi: 10.1080/0144929x.2012.671851]
36. Harris J. Altruism: Should it be Included as an Attribute of Medical Professionalism? Heal Prof Educ 2018 Mar;4(1):3-8. [doi: 10.1016/j.hpe.2017.02.005]
37. Zhang X, Liu S, Deng Z, Chen X. Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. Comput Human Behav 2017 Oct;75:797-810. [doi: 10.1016/j.chb.2017.06.028]
38. Chang HH, Chuang S. Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. Inf Manag 2011 Jan;48(1):9-18. [doi: 10.1016/j.im.2010.11.001]
39. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Adv Behav Res Ther 1978 Jan;1(4):139-161. [doi: 10.1016/0146-6402(78)90002-4]
40. Zhang X, de Pablos PO, Xu Q. Culture effects on the knowledge sharing in multi-national virtual classes: A mixed method. Comput Human Behav 2014 Feb;31:491-498. [doi: 10.1016/j.chb.2013.04.021]
41. Tamjidyamcholo A, Bin Baba MS, Tamjid H, Gholipour R. Information security – Professional perceptions of knowledge-sharing intention under self-efficacy, trust, reciprocity, and shared-language. Comput Educ 2013 Oct;68:223-232. [doi: 10.1016/j.compedu.2013.05.010]

42. Mayer RC, Davis JH, Schoorman FD. An Integrative Model Of Organizational Trust. AMR 1995 Jul;20(3):709-734. [doi: 10.5465/amr.1995.9508080335]
43. Sparks BA, Browning V. The impact of online reviews on hotel booking intentions and perception of trust. Tour Manag 2011 Dec;32(6):1310-1323. [doi: 10.1016/j.tourman.2010.12.011]
44. Karahanna E, Straub DW. The psychological origins of perceived usefulness and ease-of-use. Inf Manag 1999 Apr;35(4):237-250. [doi: 10.1016/s0378-7206(98)00096-2]
45. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
46. Montgomery DC, Peck EA, Vining G. Introduction to Linear Regression Analysis. New York: Wiley-Interscience; Feb 26, 2007:323-341.
47. Fritz CO, Morris PE, Richler JJ. Effect size estimates: Current use, calculations, and interpretation. J Exp Psychol Gen 2012;141(1):2-18. [doi: 10.1037/a0024338]
48. The Lancet. Protecting Chinese doctors. The Lancet 2020 Jan;395(10218):90. [doi: 10.1016/s0140-6736(20)30003-9]
49. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
50. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public Reporting in Germany: the Content of Physician Rating Websites. Methods Inf Med 2018 Jan 19;51(02):112-120. [doi: 10.3414/me11-01-0045]
51. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online Reviews of 500 Urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013]
52. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]

Abbreviations

ANOVA: analysis of variance
VIF: variance inflation factor

Edited by G Eysenbach; submitted 17.04.19; peer-reviewed by M Duplaga, S Bidmon; comments to author 03.10.19; revised version received 17.12.19; accepted 29.03.20; published 04.06.20.

Please cite as:
Han X, Li B, Zhang T, Qu J
Factors Associated With the Actual Behavior and Intention of Rating Physicians on Physician Rating Websites: Cross-Sectional Study
J Med Internet Res 2020;22(6):e14417
URL: http://www.jmir.org/2020/6/e14417/
doi:10.2196/14417
PMID:32496198

Original Paper

Associations Between Commercial App Use and Physical Activity: Cross-Sectional Study

Jasmine Maria Petersen[1,2], BSc; Eva Kemps[3], PhD; Lucy K Lewis[1,2], PhD; Ivanka Prichard[1,2], PhD

[1]SHAPE Research Centre, Flinders University, Adelaide, Australia
[2]Caring Futures Institute, College of Nursing and Health Sciences, Flinders University, Adelaide, Australia
[3]School of Psychology, College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia

Corresponding Author:
Jasmine Maria Petersen, BSc
SHAPE Research Centre
Flinders University
GPO Box 2100
Adelaide
Australia
Phone: 61 8 82013713
Email: jasmine.petersen@flinders.edu.au

Abstract

Background: In today's society, commercial physical activity apps (eg, Fitbit and Strava) are ubiquitous and hold considerable potential to increase physical activity behavior. Many commercial physical activity apps incorporate social components, in particular app-specific communities (allowing users to interact with other app users) or the capacity to connect to existing social networking platforms (eg, Facebook or Instagram). There is a growing need to gain greater insights into whether commercial physical activity apps and specific components of these apps (social components) are beneficial in facilitating physical activity.

Objective: This study aimed to examine the relationship between the use of commercial physical activity apps and engagement in physical activity. The social components of commercial physical activity apps (app-specific communities and existing social networking platforms) were also explored. This involved isolating specific features (eg, sharing, providing, and receiving encouragement, comparisons, and competitions) of app-specific communities and existing social networking platforms that were most valuable in facilitating physical activity.

Methods: A cross-sectional web-based survey was conducted. Participants were 1432 adults (mean age 34.1 years, 1256/1432, 88.00% female) who completed measures assessing physical activity, the use of commercial physical activity apps, and engagement with app-specific communities and existing social networking platforms.

Results: Overall, 53.14% (761/1432) of the sample reported engaging with a commercial physical activity app. The most commonly used apps were Fitbit (171/761, 22.5%), Strava (130/761, 17.1%), and Garmin (102/761, 13.4%). The use of physical activity apps was significantly associated with physical activity. Notably, the use of app-specific communities and existing social networking platforms facilitated significantly greater engagement in physical activity. The features of app-specific communities that were most beneficial in promoting engagement in physical activity were providing encouragement to a partner, receiving encouragement from close friends and family, and engaging in competitions with members of public app-specific communities. In relation to existing social networking platforms, sharing physical activity posts predicted engagement in physical activity.

Conclusions: The findings indicate that app-specific communities and existing social networking platforms are components of apps that are fundamental in facilitating physical activity. They further suggest that commercial physical activity apps afford high population level reach and hold great potential to promote engagement in physical activity, an important public health consideration.

(J Med Internet Res 2020;22(6):e17152) doi:10.2196/17152

KEYWORDS

physical activity; mobile applications; social networking

Introduction

Background

Physical activity confers many health benefits, including a reduced risk of cardiovascular disease, hypertension, osteoporosis, diabetes mellitus, obesity, mental illness, and premature mortality [1-3]. Despite this, globally, 1.4 billion adults (28%) are not meeting physical activity guidelines (150 min of moderate to vigorous physical activity per week), a figure that continues to rise [4]. This highlights the need to develop scalable interventions to increase physical activity.

Physical activity mobile apps present an innovative approach to promote engagement in physical activity due to their widespread reach, accessibility, and convenience. Recently, there has been exponential growth in the availability of commercial physical activity apps (eg, Fitbit, Strava, and Garmin) [5]. However, much of the previous research examining physical activity apps has focused on apps developed by researchers as opposed to commercially available apps [6,7]. This presents a shortcoming of research to date, such that despite the accessibility and ubiquity of commercial apps, there is limited literature exploring their use and influence on physical activity. This indicates the need to gain greater insight into the use of commercial apps to ascertain their capacity to increase levels of physical activity and, thus, improve public health.

A growing body of research examining the content of commercial apps has identified that social features are an increasingly ubiquitous component [6,8-10]. That is, many commercial apps incorporate app-specific communities, allowing users to interact with other app users by sharing physical activity data, receiving or providing encouragement (eg, likes and comments), and engaging in competitions or comparisons [8]. However, to date, there has been little examination of app-specific communities, and, in particular, the association between engagement with the features of these communities (eg, sharing and competitions) and physical activity. Insights into app-specific communities is important to ascertain their value in promoting engagement in physical activity and crucial for the development of future physical activity interventions.

Content analyses of commercial apps have identified that many physical activity apps also have the capacity to connect to existing social networking platforms such as Facebook or Instagram [6,8-10]. This has been suggested to be an important component of an app, given that a recent review [7] identified that the use of existing social networking platforms in conjunction with apps enhances engagement [11]. However, the review [7] also documented that this research area is in its infancy, and there is a need to gain greater insights into how to optimally harness existing social networking platforms in conjunction with physical activity apps. This requires identifying the features of existing social networking platforms (eg, social interaction and comparisons) that are associated with app engagement and, thus, physical activity.

Objectives

To our knowledge, no previous study has comprehensively examined commercial physical activity apps and specifically, how the social components of these apps (app-specific communities or existing social networking platforms) may be associated with physical activity. This is important given the increasing prevalence of commercial physical activity apps together with the need to isolate components of apps that are linked to physical activity engagement. Thus, the first aim of this study was to gain a comprehensive understanding of the use of commercial physical activity apps and their relationship with physical activity. The second aim was to explore the value of app-specific communities and existing social networking platforms in facilitating physical activity. More specifically, we sought to ascertain the features of app-specific communities and existing social networking platforms that were used and how these were associated with frequency of app use and engagement in physical activity.

Methods

Study Design and Participants

A web-based cross-sectional survey was conducted. Participants were recruited via the Discipline of Psychology's web-based research participation system, paid Facebook advertising, and free advertisements placed on social networking platforms (eg, Facebook, Instagram, and Twitter) for a study on *Physical Activity and Online Social Networking*. Ethical approval was obtained from the University Social and Behavioral Research Ethics Committee (protocol no. 8232). All participants provided informed consent electronically. Participants were adults, ≥18 years, and proficient in English.

Procedure

Participants completed a web-based survey through the Qualtrics platform between February and April 2019. The survey took approximately 30 min to complete and incorporated the measures listed below in the order of presentation. As a token of appreciation, participants could enter a raffle to win 1 of 5 AUD $25 (USD $15) shopping gift vouchers.

Measures

Demographics

Participants were invited to report their age, gender identity, and ethnicity.

Regular Structured Physical Activity

Regular structured physical activity was assessed following the methods of Prichard and Tiggemann [12]. Participants were invited to self-report the type, duration, and frequency of structured physical activity or sports they generally engaged in on a weekly basis. The total number of minutes of physical activity engaged in per week was then calculated by multiplying each activity's frequency by its duration. Separate physical activity totals were calculated according to the type of physical activity listed, specifically individual physical activities (eg, walking or running), gym-based activities (eg, gym classes), or sports-based activities (eg, netball or football).

Current Physical Activity App Use

Participants were asked to self-report their current use of physical activity apps, defined as apps that have the capacity to track or monitor physical activity (eg, steps or distance) or provide guided training or workouts. In particular, participants were asked to self-report using an open-ended response format, the name of the physical activity app they were currently using most frequently (main physical activity app, eg, Strava), the physical activity or sport they were using the app for, and their level of engagement with the app (number of times used per week). The apps were categorized according to their capabilities, including tracking, providing guided workouts, tracking plus providing guided workouts, or other (eg, scheduling gym classes or immersive games). The types of physical activity the apps were used for were classified as all daily activities, individual activities (eg, running, cycling, or walking), group-based activities (eg, netball, soccer, or football), gym-based activities (eg, fitness classes or personal training), or a combination of different activities (individual, group-based and gym-based activities).

Engagement With an App-Specific Community

In relation to the main physical activity app participants were currently using, they were asked to self-report their engagement with the features of the app-specific community. This included specifying how frequently on a 6-point Likert scale ranging from 0 (*never*) to 5 (*very often*) they engaged with specific features of the app community, such as sharing physical activity posts, liking and/or providing positive comments on others' posts, receiving likes and/or positive comments, comparing their physical activity performance with others, and engaging in competitions. Furthermore, participants were asked to indicate the frequency that they engaged with the aforementioned features with specific members of their app community, including partners, family, close friends, peers, public app-specific community members, and work colleagues. Example items include *Within the main physical activity app you are currently using, how often do you share posts relating to your physical activity performance with a partner?* and *Within the main physical activity app you are currently using, how often do you Like/Kudos/Cheer and/or provide positive comments on physical activity posts from close friends?* As the networks within each feature were highly correlated (α=.74), a composite score was calculated for the use of each specific feature of the app community (eg, sharing) across the different networks (eg, peers and family) while also examining the independent influence of engaging with specific networks in relation to each feature.

Participants who specified that the main physical activity app they were using incorporated an app-specific community but reported that they did not engage with it were provided with an open-ended question to determine the underlying rationale for this. Preliminary themes were established by the first author, and the responses were subsequently categorized by 2 independent coders.

Engagement With Existing Social Networking Platforms in Relation to Physical Activity

Participants were also asked to self-report their physical activity-related use of existing social networking platforms on measures developed for this study. Specifically, participants were asked to specify on a 6-point Likert scale how frequently (from *never*=0 to *very often*=5) they share physical activity posts, like and/or provide positive comments on others' posts, receive likes and/or positive comments, and compare their physical activity performance with others' physical activity posts on Facebook, Instagram, and Twitter (plus the option to specify other platform(s)). Example items included *How often do you share physical activity posts on the following social networking platforms?* and *How often do you like and/or provide positive comments on physical activity posts from other people on the following social networking platforms?* A composite score was calculated for the use of each specific feature (eg, sharing) across the different social networking platforms (eg, Facebook and Instagram).

Statistical Analysis

Data were analyzed using Statistical Package for the Social Sciences version 25 (IBM, Corp). Significance for all analyses was set at *P*<.05 (2-tailed). Overall, the study variables (with the exception of app engagement) did not deviate substantially from normality based on skewness, kurtosis, or histogram examination. Therefore, parametric tests were used for all analyses, except those that included the variable app engagement for which a nonparametric test (Kruskal-Wallis test) was used.

Descriptive statistics were used to generate demographic information. A series of independent samples *t* tests and chi-square analyses were conducted to determine differences between app users and nonusers in age, gender identity, ethnicity, and minutes of physical activity per week. Chi-square analyses were used to identify differences in app use (ie, the most commonly used apps, the capabilities of the apps used, and the activity app is used for) based on demographics (age and gender identity). A Kruskal-Wallis test with pairwise comparisons using the Dunn-Bonferroni correction was conducted to examine the relationship between app engagement (frequency of app use per week) and physical activity.

Kruskal-Wallis tests were also conducted to examine the relationships between the use of specific features of app-specific communities and app engagement among app users. In addition, 1-way analyses of variance were performed to determine differences in engagement with features of app-specific communities based on age, capabilities of the app used, and activity the app was used for, but not for gender identity (because of the small proportion of men). The aforementioned analyses were repeated using specific features of existing social networking platforms. Independent samples *t* tests and chi-square analyses were also used to determine differences between users and nonusers of the app-specific communities and existing social networking platforms.

Finally, a multiple linear regression was conducted to explore the predictors of physical activity among app users. The regression model incorporated the frequency of app usage and all features of both app-specific communities (including specific networks) and existing social networking platforms. Demographic characteristics (age, gender identity, and ethnicity) were incorporated as control variables.

Results

Sample

In total, 1640 individuals began the survey, 208 of whom did not complete it (a response rate of 87.3%), resulting in a final sample of 1432 participants. The sample had a mean age of 34.1 years (range 18-83 years) and comprised predominately female participants (1256/1432, 88.00%). Overall, the sample engaged in high levels of structured physical activity (mean 266.8 min per week, SD 219.8), and 53.14% (761/1432) reported currently engaging with a physical activity app. Table 1 presents the demographic characteristics of app users and nonusers. There were no significant differences between the 2 groups (app users and nonusers) in relation to age, gender identity, or ethnicity. However, app users engaged in significantly more structured physical activity per week than nonusers (Table 1). Overall, among those who reported engaging in physical activity, participants predominately engaged in individual physical activities (eg, walking or running, 619/858, 72.2%), followed by gym-based activities (eg, gym classes, 324/620, 52.2%), and sports-based activities (eg, netball or football, 88/324, 27.2%). This did not differ based on whether participants used an app. Relatedly, participants spent the most time engaging in individual activities per week (mean 133.6 min, SD 176.5), followed by sports-based activities (mean 113.5 min, SD 110.7), and gym-based activities (mean 83.2 min, SD 94.9). Again, this was consistent across app users and nonusers (Table 1).

Table 1. Sample characteristics of physical activity app users and nonusers (n=1432).

Characteristic	App users (n=761)	Nonusers (n=671)	*P* value[a]	Effect size, Φ	Effect size, Cohen *d*
Age (years), n (%)			.42	0.12	N/A[b]
18-25	243 (32.0)	257 (38.3)			
>25-30	115 (15.0)	76 (11.4)			
>30-40	190 (25.0)	119 (17.7)			
>40	208 (27.3)	214 (31.9)			
Gender identity, n (%)			.94	0.006	N/A
Female	668 (88.0)	588 (88.0)			
Male	84 (11.0)	73 (11.0)			
Ethnicity, n (%)			.67	0.05	N/A
White	682 (89.6)	581 (86.6)			
Asian	35 (4.6)	37 (5.5)			
Indian	10 (1.3)	13 (1.9)			
Other	34 (4.5)	40 (6.0)			
Structured physical activity (min per week), mean (SD)					
Overall structured physical activity	309.0 (214)	219.0 (216)	<.001	N/A	0.42
Individual activities (n=858)	141.9 (179.4)	120.6 (171.3)	.21	N/A	0.08
Sport-based activities (n=324)	115.5 (120.4)	110.7 (96.4)	.61	N/A	0.06
Gym-based activities (n=620)	77.4 (79.8)	92.8 (115.3)	.23	N/A	0.09

[a] Statistical significance is represented by *P*<.05.

[b] N/A: not applicable.

Physical Activity App Use

Multimedia Appendix 1 presents the physical activity apps that were most commonly used. Fitbit (171/761, 22.5%), followed by Strava (130/761, 17.1%) and Garmin (102/761, 13.4%) were the most popular apps, and this did not differ by age or gender identity. Participants most commonly engaged with apps that had the capacity to exclusively track behaviors and predominately used apps for individual activities (eg, running or walking). This was consistent across age and gender identities.

The greatest proportion of participants reported using their physical activity app on 7 occasions per week (296/761, 39.0%), followed by use on 3 occasions (102/761, 13.4%) and more than 7 occasions (70/761, 9.2%) per week. A Kruskal-Wallis Test comparing weekly physical activity duration revealed a statistically significant difference (*P*=.006) across levels of app usage. Specifically, pairwise comparisons identified that participants who used the app on 6 occasions per week engaged in significantly higher levels of structured physical activity (median 491.9 min) than those who used the app on 2 occasions per week (median 297.4 min; *P*=.003). Overall, participants who used an app 6 times per week engaged in the highest levels of structured physical activity.

Use of Social Components of Physical Activity Apps

Among app users, 3.4% (26/761) used app-specific communities exclusively, 59.9% (456/761) used existing social networking platforms exclusively, and 22.0% (167/761) used both app-specific communities and existing social networking platforms. This did not differ significantly by age, gender identity, ethnicity, capabilities of the app used (eg, tracking), the type of physical activity the app was used for, or frequency of app usage per week.

Physical Activity App Use in Conjunction With an App-Specific Community

Among app users, 59.0% (447/761) reported that the physical activity app they were currently using incorporated an app-specific community. Of these, 43.1% (193/447) reported engaging with the community. Participants who reported engaging with the app-specific communities predominantly used Strava (80/193, 41.5%), Fitbit (40/193, 20.7%), and Garmin (13/193, 6.7%). Table 2 shows that the distribution of age was significantly different between app community users and nonusers. Specifically, app community users were predominately >30 years. App community users also engaged in significantly more structured physical activity per week than nonusers (t_{445}=2.62; *P*=.009; *d*=0.25). However, there were no significant differences between users and nonusers of the app-specific community in relation to gender identity, ethnicity, capabilities of the app used (eg, tracking), the type of physical activity the app was used for, or app usage per week. Among participants who reported not engaging with the app-specific community (254/447, 57.0%), the reasons identified were privacy or security concerns, negative attitudes toward the use of the community, considered unnecessary, lack of support, beliefs regarding the nature of physical activity, disinterest in others' physical activity performance, and use of an alternative social network. Of these, the most commonly cited reasons were that the use of the community was considered unnecessary (83/249, 33.4%), disinterest in others' physical activity performance (50/246, 20.3%), and privacy or security concerns (43/246, 17.5%).

Among participants who engaged with the app-specific community (n=193), users most frequently used features that allowed the sharing of physical activity performance, providing encouragement to others' physical activity posts (eg, likes or positive comments), and receiving encouragement on one's own posts. These features were most frequently reported to be used with networks that were close friends or peers. There were no significant differences in engagement with features of app-specific communities across age, levels of app usage or according to capabilities of the app used, or the activity the app was used for.

Table 2. Sample characteristics of app-specific community users and nonusers (n=447).

Characteristic	Community users (n=193)	Nonusers (n=254)	*P* value[a]	Effect size, Φ	Effect size, *Cohen d*
Age (years), n (%)			.005	0.16	N/A[b]
18-25	34 (17.7)	77 (30.4)			
>25-30	29 (15.1)	46 (18.2)			
>30-40	59 (30.6)	65 (25.7)			
>40	70 (36.2)	65 (25.7)			
Gender identity, n (%)			.05	0.09	N/A
Female	159 (82.4)	229 (90.2)			
Male	29 (15.0)	24 (9.4)			
Ethnicity, n (%)			.45	0.11	N/A
White	177 (91.7)	231 (90.9)			
Asian	9 (4.7)	9 (3.5)			
Indian	2 (1.0)	4 (1.6)			
Other	5 (2.6)	10 (4.0)			
Structured physical activity (min per week), mean (SD)	357.6 (217.7)	305.4 (196.3)	.009	N/A	0.25
Type of app, n (%)			.26	0.09	N/A
Tracking	184 (95.3)	234 (92.1)			
Guided workouts	5 (2.6)	14 (5.5)			
Tracking and workouts	3 (1.6)	2 (0.8)			
Other (booking classes or immersive games)	1 (0.5)	4 (1.6)			
Physical activity app is used for, n (%)			.23	0.11	N/A
All daily activity	18 (9.3)	39 (15.3)			
Individual activities	154 (80.0)	184 (72.4)			
Group-based activities	0 (0.0)	1 (0.4)			
Gym-based activities	10 (5.2)	18 (7.1)			
Combination of individual and group-based and gym-based activities	6 (3.1)	6 (2.4)			

[a] Statistical significance is represented by *P*<.05.

[b] N/A: not applicable.

Physical Activity App Use in Conjunction With Existing Social Networking Platforms

Among app users, 82.0% (624/761) reported using existing social networking platforms (Facebook, Instagram, or Twitter) in relation to physical activity. There were no significant differences between users and nonusers of existing social networking platforms in relation to age, gender identity, or ethnicity. Participants who used existing social networking platforms engaged in significantly more structured physical activity than those who did not (t_{672}=2.9; *P*=.004; *d*=0.44). The features of existing social networking platforms that were most frequently used were providing encouragement on others' physical activity posts, followed by receiving encouragement on one's own physical activity posts (eg, likes or comments). Notably, there were significant differences in the frequency of engagement with features of existing social networking platforms based on age, both in terms of sharing physical activity posts ($F_{3,666}$=5.37; *P*=.001) and engaging in comparisons ($F_{3,666}$=19.0; *P*<.001). Specifically, participants aged 18-25 years shared posts to existing social networking platforms significantly less frequently than all other age groups. In addition, participants >40 years made significantly fewer comparisons relative to all other age groups. However, there were no significant differences in the frequency of engagement with features of existing social networking platforms across the frequency of app usage or according to the capabilities of the app used, or the activity the app was used for.

Exploring Predictors of Physical Activity

The regression model accounted for 42.6% of the variance in structured physical activity (R^2=0.426) and was significant ($F_{38,96}$=1.87; *P*<.01). The following variables were significant positive predictors of structured physical activity: frequency of app use (β=.25; *P*=.009), providing encouragement to a partner (β=.52; *P*=.005), receiving encouragement from close friends (β=.59; *P*=.01) and family (β=.48; *P*=.02), engaging in competitions with members of a public app-specific community (β=.38; *P*=.001), and sharing posts to existing social networking platforms (β=.31; *P*=.004). In addition, the following variables were significant negative predictors of structured physical activity: sharing physical activity posts with a partner (β=−.40; *P*=.007), providing encouragement to close friends (β=−.57; *P*=.01), receiving encouragement from members of a public app-specific community (β=−.35; *P*=.04), and engaging in competitions with a partner (β=−.30; *P*=.04, Table 3).

Table 3. Multiple regression analysis examining predictors of structured physical activity among app users (n=761).

Variable	β	*t* test	*P* value[a]
Gender identity	−.05	−0.62	.53
Age	−.02	−0.24	.81
Ethnicity	-.05	-0.48	.63
App engagement (frequency)	.25	2.64	.009
App-specific communities			
Sharing posts			
Partner	−.40	−2.74	.007
Family	−.18	−1.09	.28
Close friends	.08	0.50	.62
Peers	.02	0.14	.88
Public app community	−.02	−0.16	.87
Colleagues	−.15	−1.24	.21
Providing encouragement			
Partner	.52	2.90	.005
Family	−.08	−0.45	.64
Close friends	−.57	−2.50	.01
Peers	.19	0.96	.34
Public app community	.29	1.97	.05
Colleagues	.10	0.53	.59
Receiving encouragement			
Partner	−.14	−0.84	.40
Family	.48	2.28	.02
Close friends	.59	2.41	.01
Peers	.01	0.07	.94
Public app community	−.35	−1.98	.04
Colleagues	−.15	−0.81	.41
Engagement in competitions			
Partner	−.30	−2.0	.04
Family	−.06	−0.40	.69
Close friends	−.16	−1.16	.24
Peers	−.18	−1.25	.21
Public app community	.38	3.27	.001
Colleagues	−.07	−0.55	.57
Engagement in comparisons			
Partner	.15	0.93	.35
Family	−.04	−0.24	.81
Close friends	−.17	−1.05	.29
Peers	.15	0.93	.35
Public app community	.01	0.06	.94
Colleagues	.06	0.35	.72
Existing social networking platforms			
Sharing posts	.31	2.92	.004

Variable	β	*t* test	*P* value[a]
Providing encouragement	−.07	−0.63	.52
Receiving encouragement	.16	1.43	.15
Engagement in comparisons	.02	0.20	.83

[a] Statistical significance is represented by $P<.05$.

Discussion

Principal Findings

This study aimed to provide a comprehensive examination of the use of commercial physical activity apps and the relationship between app usage and physical activity. In addition, we sought to explore the use of social components of apps (app-specific communities and existing social networking platforms) and their value in facilitating engagement in physical activity. This study is timely, given the ubiquity of commercial physical activity apps coupled with the need to understand how specific components of these apps may be beneficial in facilitating physical activity.

Overall, the findings demonstrate that the use of physical activity apps is common, with over half of the participants reporting that they currently use a physical activity app. Our findings are consistent with Krebs and Duncan [13], who reported that in a large, diverse US sample (50% female), 58.2% had downloaded a health-related app, of which 52.8% used the app to track physical activity. This reflects both the omnipresence of commercial physical activity apps and their capacity to have high-population level reach. Notably, physical activity app users engaged in significantly more structured physical activity, consistent with findings from a previous study documenting that physical activity app users were 27% more likely to engage in physical activity than nonusers [14]. In this study, app users predominantly used Fitbit, Strava, and Garmin, with the primary function of these apps being to track or monitor behavior. This may explain the higher levels of physical activity among app users, given that self-monitoring is a behavior change technique consistently associated with increased physical activity [15,16]. These findings suggest that commercial physical activity apps may have great potential to influence physical activity. However, it must be acknowledged that the causality of the relationship between app use and physical activity is presently unclear, in that those who engage in high levels of physical activity may be attracted to apps to monitor their behavior. Longitudinal research examining app use and physical activity over time is needed to ascertain the direction of this relationship.

This study identified that participants most commonly engaged with physical activity apps on 7 occasions per week, and in line with previous research, the frequency of app usage was significantly associated with physical activity [11]. Interestingly, participants who used the apps on 6 occasions per week engaged in the highest duration (minutes) of structured physical activity per week. This indicates that relatively high app use is associated with high levels of physical activity, reflective of the previously cited dose-response relationship between app use and behavioral outcomes (eg, physical activity) [17]. The present findings further highlight that app usage is an important consideration in appropriately leveraging apps to promote engagement in physical activity. This emphasizes the importance of examining specific components of apps that may be harnessed to increase app usage, namely social components.

A novel aspect of this study was its comprehensive examination of the social components (app-specific communities and existing social networking platforms) of commercial physical activity apps. Interestingly, the use of the social components of apps differed markedly, such that most app users engaged exclusively with existing social networking platforms (in relation to physical activity, 456/761, 59.9%), whereas far fewer engaged with both app-specific communities and existing social networking platforms (167/761, 22.0%) or app-specific communities exclusively (26/761, 3.4%). This is perhaps not surprising given that existing social networking platforms are immensely popular, afford widespread reach, and achieve high levels of sustained engagement [18]. Age is another factor that may have contributed to the difference in the use of the social components of apps, such that users of app-specific communities were predominately >30 years (66.8%); by contrast, age was not associated with the use of existing social networking platforms. This suggests that app-specific communities are most appropriate for a specific subgroup of the population (>30 years), whereas existing social networking platforms could be harnessed for the population more broadly. Relatedly, future research could consider examining other factors that may influence the use of app-specific communities and/or existing social networking platforms, such as one’s presence on social networking platforms (ie, those with vs without an established social networking presence).

Despite the differences in the usage of app-specific communities and existing social networking platforms, the social features across both were used similarly, with providing and receiving encouragement the most popular features. Interestingly, the findings also indicate that demographic characteristics may be linked to the likelihood that individuals will utilize specific features. For example, in relation to existing social networking platforms, those aged 18-25 years shared posts less frequently than all other age groups, whereas those aged >40 years engaged in comparisons less frequently relative to all other age groups. This is consistent with a recent review [19] of mobile health interventions documenting that preferences for social features that facilitated comparisons, competitions, or social support varied among participants and purported that this may be linked to individual differences (eg, competitiveness). When leveraging the features of social components of physical activity apps (eg, existing social networking platforms), a 1-size-fits-all approach is not appropriate, and individual differences must be considered. Future research may usefully extend this understanding by examining whether other individual difference factors such as psychological characteristics (competitiveness and social comparisons) may influence the use of specific features.

Notably, the use of app-specific communities and existing social networking platforms was associated with significantly higher engagement in structured physical activity. This is a novel finding given that it suggests that the social components of apps may play a fundamental role in facilitating engagement in physical activity. This may be attributed to the unique capacity of app-specific communities and existing social networking platforms to generate social support [20], an important determinant of engagement in physical activity [21,22]. Another explanation is that individuals who engage in more physical activity have more content to share and are more likely to engage with the social components of apps, which, in turn, may foster supportive interactions. This study demonstrates the value of social components of apps in facilitating physical activity and the need to further examine app-specific communities and existing social networking platforms in future research.

Finally, the regression analysis revealed that the frequency of app usage was a significant predictor of structured physical activity, indicating the need to determine strategies that will facilitate app use. Interestingly, in relation to app-specific communities, providing encouragement to a partner, receiving encouragement from close friends and family, and engagement in competitions with members of public app-specific communities were positive predictors of physical activity. Conversely, sharing physical activity posts and engaging in competitions with a partner, providing encouragement to close friends, and receiving encouragement from members of public app-specific communities were negative predictors of physical activity. These findings indicate that receiving encouragement from strong ties (close friends and family) is beneficial in facilitating engagement in physical activity, whereas receiving encouragement from weak ties (public app-specific communities) is not. This is consistent with previous research showing that strong ties provide emotional support (encouragement, empathy) [23,24], which is linked to improvements in health outcomes [25], and an increased likelihood that one will initiate and maintain engagement in physical activity [26]. Conversely, weak ties often only provide informational support (advice or suggestions) [23,24], shown to be negatively associated with health behavior [27], and this has been attributed to receiving information or advice that is unwanted or in surplus [28]. The findings do, however, suggest that engaging in competitions with weak ties (public app-specific communities) is advantageous in promoting physical activity, whereas engaging in competitions or behaviors that may generate competitions (sharing physical activity posts) with strong ties (eg, partner) is negatively associated with physical activity. This is in line with existing research documenting that comparisons (generating competitions) with strong ties elicits greater pressure and fear of experiencing shame and embarrassment, ultimately decreasing the likelihood that one will engage in the behavior [29,30]. Together, these findings provide an important understanding of the specific features (and networks) of app-specific communities that are most beneficial in facilitating physical activity, and thus should be leveraged in future app-based interventions.

The regression analysis also showed that in relation to existing social networking platforms, sharing physical activity posts positively predicted engagement in physical activity. This is perhaps not surprising given that a recent study [27] found that sharing posts related to tracked health information (eg, physical activity, sleep, or calories) to existing social networking platforms is positively associated with social support, and this, in turn, predicts engagement in the associated health behavior. This highlights the need for future research to further explore the use of existing social networking platforms in relation to physical activity apps and physical activity behavior. In addition, future studies could explore how different social networking environments (eg, network size or composition) may interact with apps to influence physical activity.

Implications

Our findings have important implications for informing the design of future app-based interventions. They demonstrate that commercial physical activity apps, in particular, those that facilitate self-monitoring (eg, Fitbit, Strava, and Garmin) hold great potential to promote engagement in physical activity. The convenience, accessibility, and affordability of commercial apps coupled with their capacity to facilitate physical activity highlight that future app-based interventions should harness commercial apps, as opposed to previous interventions that have predominately incorporated researcher-developed apps [7]. The findings also indicate that the social components of apps are important in promoting physical activity and, thus, fundamental in the development of effective app-based interventions. More specifically, in relation to app-specific communities, receiving encouragement from close friends and family, providing encouragement to a partner, and engagement in competitions with members of public app-specific communities were shown to be the most beneficial features in facilitating physical activity and, thus, should be leveraged to maximize effectiveness. However, relatively few app users engaged with app-specific communities (193/447, 43.1%), with commonly cited barriers to using the communities, including disinterest and privacy or security concerns. Nevertheless, app-specific communities show great potential in facilitating physical activity, and thus, these barriers must be considered and overcome in the design of future apps and app-based interventions. The findings also suggest that existing social networking platforms are commonly used in relation to physical activity, and notably, sharing physical activity posts to these platforms predicts engagement in physical activity. Thus, existing social networking platforms will be an important component of future app-based interventions given their capacity to achieve high levels of use and promote engagement in physical activity.

This study also has important implications for guiding future research. Experimental evidence is now needed to isolate the influence of the social components of apps and their associated features on physical activity. Future research should also endeavor to ascertain the mechanisms (eg, social support and self-efficacy) underlying the capacity of app-specific communities and existing social networking platforms to facilitate physical activity. Finally, longitudinal research is needed to determine the value of app-specific communities and existing social networking platforms in promoting sustained app use, and thus prolonged engagement in physical activity.

Limitations

As with all studies, there are a number of limitations that need to be acknowledged. First, the sample consisted predominantly of white women, and the participants' origin (country/region) was unknown. These 2 sample characteristics limit the generalizability of the findings to the population more broadly. Nevertheless, app usage rates were similar to those previously reported in a large, diverse US sample [13]. Second, the sample as a whole engaged in high levels of structured physical activity, indicative of self-selection bias, and, thus, may not be representative of the general population. Third, participants self-reported structured physical activity, which may have resulted in under- or overreporting. In addition, the assessment of structured physical activity did not capture incidental physical activity, which is often recorded by apps that track or monitor daily or individual physical activities. Thus, future research should consider using accelerometer-derived measures of physical activity in conjunction with self-report measures that have the capacity to capture both structured and incidental physical activity. Finally, there are a number of factors that may influence app use and/or the association between app use and physical activity such as socioeconomic status (SES), fitness device ownership, and overall engagement with the app-specific community, which were not measured in this study. Many physical activity apps are often used in conjunction with fitness devices (eg, Fitbits and Apple Watches), which may be too expensive for some individuals. As such, future research could examine potential interactions between SES, fitness device ownership, app use, and physical activity.

Conclusions

Notwithstanding these limitations, this study provides an important contribution to existing literature by comprehensively exploring the use of commercial physical activity apps and their associated social components in a large cross-sectional sample. The findings indicated that the use of commercial physical activity apps facilitates engagement in physical activity and, therefore, have great potential to disseminate scalable interventions to improve health behavior. This study also provided a nuanced insight into app-specific communities and existing social networking platforms, identifying that they are components of apps that are valuable in promoting physical activity, and should be harnessed in the development of future app-based interventions. Together, these findings highlight the importance of further examining the social components of apps and gaining an understanding of the mechanisms underlying their influence on physical activity. In so doing, this study has demonstrated that commercial physical activity apps afford high population level reach and hold great potential to facilitate engagement in physical activity. Thus, future interventions aimed at increasing physical activity should further explore commercial physical activity apps and their associated social components.

Acknowledgments

Authors would like to thank Theodora Sevastidis for coding the qualitative responses and the participants for completing the online questionnaire.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Physical activity app use of the sample (n=761).
[PDF File (Adobe PDF File), 72 KB - jmir_v22i6e17152_app1.pdf]

References

1. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Barbour KA, Edenfield TM, Blumenthal JA. Exercise as a treatment for depression and other psychiatric disorders: a review. J Cardiopulm Rehabil Prev 2007;27(6):359-367. [doi: 10.1097/01.HCR.0000300262.69645.95] [Medline: 18197069]
3. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
4. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
5. Aitken M, Clancy B, Nass D. IQVIA: The Human Data Science Company . 2017. The Growing Value of Digital Health URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-07-29]

6. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
7. Petersen JM, Prichard I, Kemps E. A comparison of physical activity mobile apps with and without existing web-based social networking platforms: systematic review. J Med Internet Res 2019 Aug 16;21(8):e12687 [FREE Full text] [doi: 10.2196/12687] [Medline: 31420956]
8. Mollee J, Middelweerd A, Kurvers R, Klein M. What technological features are used in smartphone apps that promote physical activity? A review and content analysis. Pers Ubiquit Comput 2017 Jul 5;21(4):633-643. [doi: 10.1007/s00779-017-1023-3]
9. Cowan LT, van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]
10. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
11. Schoeppe S, Alley S, van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
12. Prichard I, Tiggemann M. Relations among exercise type, self-objectification, and body image in the fitness centre environment: the role of reasons for exercise. Psychol Sport Exerc 2008 Nov;9(6):855-866. [doi: 10.1016/j.psychsport.2007.10.005]
13. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
14. Litman L, Rosen Z, Spierer D, Weinberger-Litman S, Goldschein A, Robinson J. Mobile exercise apps and increased leisure time exercise activity: a moderated mediation analysis of the role of self-efficacy and barriers. J Med Internet Res 2015 Aug 14;17(8):e195 [FREE Full text] [doi: 10.2196/jmir.4142] [Medline: 26276227]
15. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
16. Knittle K, Nurmi J, Crutzen R, Hankonen N, Beattie M, Dombrowski SU. How can interventions increase motivation for physical activity? A systematic review and meta-analysis. Health Psychol Rev 2018 Sep;12(3):211-230. [doi: 10.1080/17437199.2018.1435299] [Medline: 29385950]
17. Poirier J, Cobb NK. Social influence as a driver of engagement in a web-based health intervention. J Med Internet Res 2012 Feb 22;14(1):e36 [FREE Full text] [doi: 10.2196/jmir.1957] [Medline: 22356829]
18. Statista. 2018. Number of Monthly Active Facebook Users Worldwide as of 1st Quarter 2018 URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ [accessed 2019-07-04]
19. Tong HL, Laranjo L. The use of social features in mobile health interventions to promote physical activity: a systematic review. NPJ Digit Med 2018;1:43 [FREE Full text] [doi: 10.1038/s41746-018-0051-3] [Medline: 31304323]
20. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
21. Duncan TE, McAuley E. Social support and efficacy cognitions in exercise adherence: a latent growth curve analysis. J Behav Med 1993 Apr;16(2):199-218. [doi: 10.1007/bf00844893] [Medline: 8315646]
22. Courneya KS, McAuley E. Cognitive mediators of the social influence-exercise adherence relationship: a test of the theory of planned behavior. J Behav Med 1995 Oct;18(5):499-515. [doi: 10.1007/bf01904776] [Medline: 8847717]
23. Granovetter MS. The strength of weak ties. Am J Sociol 1973 May;78(6):1360-1380. [doi: 10.1086/225469]
24. Williams D. On and off the 'net: scales for social capital in an online era. J Comp Mediated Comm 2006 Jan;11(2):593-628. [doi: 10.1111/j.1083-6101.2006.00029.x]
25. Robinson JD, Turner JW, Tian Y, Neustadtl A, Mun SK, Levine B. The relationship between emotional and esteem social support messages and health. Health Commun 2017 Nov 28;34(2):1-7. [doi: 10.1080/10410236.2017.1405476] [Medline: 29182380]
26. Kouvonen A, de Vogli R, Stafford M, Shipley MJ, Marmot MG, Cox T, et al. Social support and the likelihood of maintaining and improving levels of physical activity: the Whitehall II study. Eur J Public Health 2012 Aug;22(4):514-518 [FREE Full text] [doi: 10.1093/eurpub/ckr091] [Medline: 21750013]
27. Oeldorf-Hirsch A, High AC, Christensen JL. Count your calories and share them: health benefits of sharing mhealth information on social networking sites. Health Commun 2019 Sep;34(10):1130-1140. [doi: 10.1080/10410236.2018.1465791] [Medline: 29683721]
28. McLaren RM, High AC. The effect of under- and over-benefited support gaps on hurt feelings, esteem, and relationships. Commun Res 2015 Sep 21;46(6):785-810. [doi: 10.1177/0093650215605155]

29. Garcia SM, Tor A, Schiff TM. The psychology of competition: a social comparison perspective. Perspect Psychol Sci 2013 Nov;8(6):634-650. [doi: 10.1177/1745691613504114] [Medline: 26173228]
30. Zhang J, Centola D. How social networks shape social comparison. In: Suls J, Collins RL, Wheeler L, editors. Social Comparison, Judgment, and Behavior. New York, USA: Oxford University Press; 2018.

Abbreviations

SES: socioeconomic status

Edited by G Eysenbach; submitted 21.11.19; peer-reviewed by Z Rosen, S King-Dowling; comments to author 23.01.20; revised version received 05.03.20; accepted 11.03.20; published 03.06.20.

Please cite as:
Petersen JM, Kemps E, Lewis LK, Prichard I
Associations Between Commercial App Use and Physical Activity: Cross-Sectional Study
J Med Internet Res 2020;22(6):e17152
URL: https://www.jmir.org/2020/6/e17152
doi:10.2196/17152
PMID:32490836

Original Paper

Effect of BMI and Perceived Importance of Health on the Health Behavior of College Students: Cross-Sectional Study

Wan-Chen Hsu[1*], PhD; Chia-Hsun Chiang[2*], PhD

[1]National Kaohsiung University of Sciences and Technology, Kaohsiung, Taiwan
[2]National Sun Yat-Sen University, Kaohsiung, Taiwan
[*]all authors contributed equally

Corresponding Author:
Chia-Hsun Chiang, PhD
National Sun Yat-Sen University
70 Lienhai Rd
Kaohsiung, 80424
Taiwan
Phone: 886 7 5252000
Email: d996050002@gmail.com

Abstract

Background: Both body mass index (BMI) and the perceived importance of health have received a lot of attention, but few studies have fully investigated the interaction of their effects on health behaviors.

Objective: This study investigates the effects of BMI and the perceived importance of health on health behaviors (patterns of eating, sleeping, and exercising) among college students in Taiwan.

Methods: A survey was conducted with 334 students to assess their perception of the importance of health (using indicators) and their health behavior (using the Health Behaviors Scale). Respondent BMI was calculated from self-reported body weight and height. Descriptive statistical analysis, independent t test analysis, two-way analysis of variance (ANOVA), and one-way ANOVA were conducted.

Results: The results showed a significant difference between genders in health behaviors among college students (eating: t_{332}=2.17, P=.03; exercise: t_{332}=5.57, P<.001; sleep: t_{332}=2.58, P=.01). Moreover, there was an interaction between BMI and perception (of the importance of health) for exercise behaviors ($F_{2,328}$=3.50, P=.03), but not for eating behaviors ($F_{2,328}$=0.12, P=.89) or sleep behaviors ($F_{2,328}$=1.64, P=.20).

Conclusions: This study establishes, for the first time, the interaction of BMI and the perceived importance of health on health behaviors. The perception of health was found to have a significant effect on exercise behaviors. Thus, the perception of health plays a significant role in the exercise behaviors of college students in Taiwan. This finding provides researchers, policy makers, and practitioners with evidence, and consequently, opportunities for focusing on preventive actions. The findings suggest that increasing the importance of health in the perception of college students, should be the focus of efforts to help students exercise more regularly.

(J Med Internet Res 2020;22(6):e17640) doi:10.2196/17640

KEYWORDS

body mass index; college students; health behavior; perceived importance of health

Introduction

Both identity and health habits are formed during a young person's transition from adolescence to adulthood [1]. The period during which one attends college is the ideal time to establish positive healthy behavior and eliminate unhealthy ones [2]. Although these years provide personal freedom and opportunities for new experiences, there is also the potential to adopt unhealthy and risky behavior [3]. Common health issues that affect young people at the start of college are poor sleep quality and lack of sleep [4]. In particular, students in the first year of college face several stressors and experience changes in their dietary and exercise patterns [5]. In its 2018 annual report, Taiwan's Health Promotion Administration highlighted the increase in the number of overweight and obese students over 18 years of age [6]. The findings of their Changes in

National Nutrition and Health Survey suggest the need for a deeper understanding of the health behaviors of Taiwanese college students [6].

Healthy behavior, such as making good dietary choices, exercising regularly, and maintaining regular sleep patterns, consists of positive actions that maintain or enhance health [7]. In general, healthy eating, sleeping, and exercise habits are each related to individual factors [8]. Scholars have argued that the perception that health is important prompts an individual to adopt a health-promoting lifestyle [9]. Previous studies have shown that individuals who place the utmost importance on health tend to adopt healthy behavior with respect to diet and nutrition, exercise [7-9], and sleep [7,8].

The link between body mass index (BMI) and health behaviors has also been the focus of research. Diet-related behaviors [10], dietary fat intake, amount of physical activity, amount of sedentary leisure time [11], and unhealthy behaviors [12] were all found to be associated with BMI. Students with a low BMI were less likely to snack and more likely to eat breakfast regularly [13]. In contrast, a high BMI was associated with unhealthy dietary patterns [14], eating irregularly [15], or eating a diet low in fiber-density [16]. Obesity and being overweight have been associated with being less likely to exercise regularly [17]. A U-shaped curvilinear association was found between sleep duration and BMI [18]—being overweight [19] or being underweight [20] was associated with poor sleep quality.

The roles of BMI and the perceived importance of health have received a lot of attention, but few studies have fully investigated the interaction of their effects on health behaviors. Because an in-depth study of these aspects would be useful in guiding the development of an effective educational program that promotes healthy behavior among college students, the aim of this study was to investigate the interaction of their effects. We hypothesize that there is an interaction between the effects of the perceived importance of health and BMI on the health behaviors of college students. Additionally, we hypothesize there is a difference between male and female health behaviors. Gender has been found to be a strong predictor of health behaviors [21] and studies have shown that female college students are more likely than male college students to have poor eating habits [22,23], poor exercise habits [23], and poor sleep habits [8].

Methods

Respondents

A questionnaire was administered to 360 college students from 12 colleges of which 334 valid questionnaires were returned (334/360, 92.8%). Of the total respondents, 50.0% were male (167/334) and 50.0% were female (167/334); 47.9% (160/334) attended college in the northern region, 23.1% (77/334) attended college in the central region, and 29.0% (97/334) attended college in the southern region of Taiwan; and the mean age was 20.70 (SD 1.35) years.

Questionnaire

Data collected through the questionnaire included age, gender, height, weight, an item indicating perceived importance of health, and the 12-item Health Behaviors Scale. Using the reference standard of Taiwan's Ministry of Health and Welfare [24], respondents were classified into three groups according to their BMI: values less than 18.5, values under 24 and greater than or equal to 18.5, and values greater than or equal to 24.

To assess respondents' perceived importance of health, we employed an item designed by Beşer et al [9]. The degree of importance assigned to health was measured by a 5-point Likert scale ranging from "none" (coded as 1) to "very much" (coded as 5) in response to the question: "To what degree do you give importance to your health?" [9]. To avoid group divisions that were not meaningful, respondents were classified into two groups according to their scores: scores less than or equal to 3 (low importance) and scores greater than or equal to 4 (high importance).

The respondents' health behaviors were evaluated using Chiang et al's [8] Health Behaviors Scale which consists of 12 items that address three aspects of health behavior: exercise (four items), eating (four items), and sleep behaviors (four items). Using item analysis, exploratory factor analysis, and confirmatory factor analysis, Chiang et al [8] showed that the Health Behaviors Scale is a valid and reliable measure of health behaviors for Taiwanese college students (Cronbach α=.83). Respondents answered the 12-item Health Behaviors Scale on a 5-point Likert scale with scores ranging from 1 (never) to 5 (always). Higher scores indicated that respondents have regular physical activity and exercise, healthy eating habits, healthy sleep habits, and good quality sleep. In terms of internal consistency, the exercise (Cronbach α=.79), eating (Cronbach α=.77), and sleep (Cronbach α=.80) showed good reliability.

Data Analysis

Statistical analysis was performed using SPSS software (version 20.0; IBM Corp). First, for a better understanding of the respondents' characteristics, descriptive statistical analysis was performed. Second, an independent *t* test (two-tailed) was used to identify if a gender difference in health behaviors existed. A *P* value less than .05 indicated that there was a statistically significant difference between genders. Finally, a two-way analysis of variance (ANOVA) was used to determine the effects of BMI and perceived importance of health on exercise, eating, and sleep behaviors among respondents. If the interaction of BMI and the perceived importance of health was significant ($P<.05$), then a one-way ANOVA was performed to assess the simple effects of the perceived importance of health and BMI.

Results

Descriptive Statistics

Table 1 presents a descriptive summary of each health behavior for the corresponding perceived importance of health and BMI groups. The mean scores were 13.46 (SD 3.06) for eating behaviors, 13.75 (SD 3.61) for exercise behaviors, and 10.66 (SD 3.56) for sleep behaviors.

Table 1. Descriptive summary for each health behavior by perceived importance of health and BMI.

Respondent groupings, n (%)		Health behavior scores		
		Eating, mean (SD)	Exercise, mean (SD)	Sleep, mean (SD)
Low perception of health importance	156 (100)	12.31 (2.98)	12.81 (3.41)	9.51 (3.21)
BMI<18.5	21 (13.5)	11.38 (2.78)	11.86 (3.77)	8.05 (2.29)
18.5≤BMI<24	88 (56.4)	12.32 (3.17)	13.25 (3.49)	9.57 (3.35)
BMI≥24	47 (30.1)	12.72 (2.63)	12.43 (3.01)	10.06 (3.15)
High perception of health importance	178 (100)	14.47 (2.78)	14.57 (3.60)	11.67 (3.55)
BMI<18.5	20 (11.2)	13.75 (2.10)	11.90 (2.40)	11.70 (2.96)
18.5≤BMI<24	121 (68.0)	14.48 (2.66)	14.60 (3.62)	11.76 (3.69)
BMI≥24	37 (20.8)	14.84 (3.40)	15.89 (3.32)	11.38 (3.43)
All	334 (100)	13.46 (3.06)	13.75 (3.61)	10.66 (3.56)
BMI<18.5	41 (12.3)	12.54 (2.72)	11.88 (3.14)	9.83 (3.19)
18.5≤BMI<24	209 (62.6)	13.57 (3.07)	14.03 (3.62)	10.84 (3.70)
BMI≥24	84 (25.1)	13.65 (3.15)	13.95 (3.58)	10.64 (3.32)

Gender Differences in Health Behaviors

Table 2 reveals gender differences in college students’ eating, exercise, and sleep behaviors indicating that male students have healthier eating behaviors (*P*=.03), better exercise habits (*P*<.001), and better sleep quality (*P*=.01) than female students.

Table 2. Health behavior scores showing gender differences.

Health behaviors	Gender, mean (SD)		*t* test (*df*)	*P* value
	Male (n=167)	Female (n=167)		
Eating	13.83 (3.10)	13.10 (2.99)	2.17 (332)	.03
Exercise	14.80 (3.69)	12.69 (3.21)	5.57 (332)	<.001
Sleep	11.16 (3.51)	10.17 (3.54)	2.58 (332)	.01

Effects of BMI and Perceived Importance of Health on Eating, Exercise, and Sleep Behaviors

As shown in Table 3, for eating behaviors, there were statistically significant main effects for BMI ($F_{2,328}$=3.66, *P*=.03) and for the perceived importance of health ($F_{1,328}$=29.44, *P*<.001); however, the interaction effect between BMI and the perceived importance of health was not statistically significant ($F_{2,328}$=.12, *P*=.89).

For exercise behaviors, there were statistically significant main effects for BMI ($F_{2,328}$=6.91, *P*=.001) and for the perceived importance of health ($F_{1,328}$=14.66, *P*<.001) as well as a statistically significant interaction between BMI and the perceived importance of health ($F_{2,328}$=3.50, *P*=.03).

For sleep behaviors, there was a statistically significant main effect for the perceived importance of health ($F_{1,328}$=26.99, *P*<.001); however, there was no statistically significant main effect for either BMI ($F_{2,328}$=1.03, *P*=.36) or the interaction between BMI and the perceived importance of health ($F_{2,328}$=1.64, *P*=.20).

As indicated by the simple effects analysis of the perceived importance of health and BMI on exercise behaviors shown in Table 4, respondents who considered health to be important and whose BMI was greater than or equal to 18.5 had significantly higher exercise behavior scores than respondents who did not consider health to be as important and whose BMI was greater than or equal to 18.5. In addition, respondents whose BMI was greater than or equal to 18.5 and who considered health to be important had significantly higher exercise behavior scores than respondents whose BMI was less than 18.5 and who considered health to be important.

Table 3. Statistical results of the two-way analysis of variance for the effects of perceived importance of health and BMI on health behaviors.

Health behaviors	BMI		Importance of health		BMI × Importance of health	
	F test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
Eating	3.66 (2,328)	.03	29.44 (1,328)	<.001	0.12 (2,328)	.89
Exercise	6.91 (2,328)	.001	14.66 (1,328)	<.001	3.50 (2,328)	.03
Sleep	1.03 (2,328)	.36	26.99 (1,328)	<.001	1.64 (2,328)	.20

Table 4. Simple effects of the perceived importance of health and BMI on exercise behaviors.

Exercise behaviors scores, mean (SD)	Perception groups		Perception effect	
	Low	High	*F* test (*df*)	*P* value
BMI groups				
BMI<18.5	11.86 (3.77)	11.90 (2.40)[a,b]	0.002 (1,39)	.97
18.5≤BMI<24	13.25 (3.49)	14.60 (3.62)[a]	7.34 (1,207)	.007
BMI≥24	12.43 (3.01)	15.89 (3.32)[b]	25.03 (1,82)	<.001
BMI effect				
F test (*df*)	1.87 (2,153)	8.72 (2,175)	N/A[c]	N/A[c]
P value	.16	<.001	N/A[c]	N/A[c]

[a] Indicates belonging to the a pair, and statistically significant difference between the pairings (*P*=.006)

[b] Indicates belonging to the b pair, and statistically significant difference between the pairings (*P*<.001).

[c] N/A: Not applicable.

Discussion

Principal Findings

This study aimed to identify the interaction the effects of BMI and perceived importance of health on eating, exercise, and sleep behaviors. Moreover, there were significant gender differences in health behaviors among Taiwanese college students. Finally, there was an interaction between the effects of BMI and perceived importance of health for exercise behaviors, but not for eating or sleep behaviors. The effect of the perceived importance of health on exercise behavior was only true for those with BMI greater than 18.5; for underweight individuals, the perceived importance of health had no effect on exercise behaviors. Thus, the research hypothesis was partly supported.

Previous studies have found that being overweight is associated with physical inactivity and sedentary daily habits [17]. Thus, overweight and obese students may require encouragement to undertake physical activity. This study found an interaction between the effects of BMI and the perceived importance of health on exercise behaviors. College students with normal and above normal BMI who placed importance on health were more likely to exercise regularly than those with below normal BMI. Consistent with previous studies [7-9], we found the perceived importance of health plays a significant role in adopting healthy exercise behavior. Therefore, increasing the perception of health among overweight students may encourage them to participate more in physical activities.

This study found that the mean score of sleep behaviors was 10.66 (range 4-20). Previous studies have shown that college students have poor sleep quality and irregular sleeping habits [8,25,26]. College students with higher stress levels experience poor sleep quality [25,26]. Many health education programs are focused on improving knowledge and skills related to regular exercise and a healthy diet. However, sleep quality and problems related to sleep are overlooked [27]. School and government authorities need to be more proactive in designing appropriate sleep and stress management strategies to help college students improve their sleep quality and maintain positive sleep habits.

This study showed that male students have healthier eating behaviors, better exercise habits, and better sleep quality than female students supporting the findings of previous studies [8,22,23]. These results suggest that gender-specific eating, exercise, and sleep intervention programs for college students are necessary.

Finally, this study found that the interaction between the effects of BMI and perceived importance of health did not affect health behaviors related to eating and sleeping; however, the study also revealed that the eating behaviors of college students with different BMI show significant differences but sleep behaviors do not. College students with different levels of perceived importance of health also have significant differences in their eating and sleeping behaviors. These findings reveal that college students' eating behaviors are influenced by their BMI and both eating and sleeping behaviors are influenced by perceived importance of health.

Limitations

The study sample consisted of respondents in a higher education setting and was restricted by age to college students and by place to Taiwan. Consequently, the findings should not be overgeneralized and must be interpreted with consideration of the sample's homogeneity. Finally, the perceived importance of health was measured using only one item. Future studies should include respondents of different ages as well as develop further instruments to measure this construct (the perceived importance of health).

Conclusions

This study found that the interaction between the effects of BMI and the perceived importance of health does not affect the health behaviors of eating and sleeping. Previous studies have shown that college students' choice to major in medicine and their self-rated health were related to their health behaviors [7,8]. Future studies should consider the effects of the interactions of these individual factors as well as BMI on eating and sleeping behaviors, to investigate how best to help students adopt regular eating habits and improve their sleep quality.

This study established, for the first time, the interaction between the effects of BMI and the perceived importance of health on health behaviors. This finding provides researchers, policy makers, and practitioners in the field with evidence and opportunities for focusing on preventive action. The findings of the study suggest that increasing the perceived importance of health should guide efforts to help students adopt better exercise habits. There were also significant gender differences in health behaviors among college students suggesting that health education practitioners should design gender-specific health behavior intervention programs for college students.

Conflicts of Interest

None declared.

References

1. Wong CA, Greeno J, Perrin E, Peterson E, Chung R, Skinner A. Young and not so invincible: health behaviors and patient-reported health outcomes among adolescents and young adults. J Adolesc Health 2018 Feb;62(2):S132. [doi: 10.1016/j.jadohealth.2017.11.269]
2. Stanziano DC, Butler-Ajibade P. Differences in health-related behaviors and body mass index risk categories in African American women in college. J Natl Med Assoc 2011 Jan;103(1):4-8 [FREE Full text] [doi: 10.1016/s0027-9684(15)30236-4] [Medline: 21329240]
3. Scott-Sheldon LA, Carey KB, Carey MP. Health behavior and college students: does Greek affiliation matter? J Behav Med 2008 Feb;31(1):61-70 [FREE Full text] [doi: 10.1007/s10865-007-9136-1] [Medline: 17999173]
4. Buboltz WC, Brown F, Soper B. Sleep habits and patterns of college students: a preliminary study. J Am Coll Health 2001 Nov;50(3):131-135. [doi: 10.1080/07448480109596017] [Medline: 11765249]
5. Hoffman DJ, Policastro P, Quick V, Lee SK. Changes in body weight and fat mass of men and women in the first year of college: a study of the "freshman 15". J Am Coll Health 2006;55(1):41-45. [doi: 10.3200/JACH.55.1.41-46] [Medline: 16889314]
6. Taiwan Health Promotion Administration. Annual Report. 2018. URL: https://www.hpa.gov.tw/Pages/ashx/File.ashx?FilePath=~/File/Attach/10231/File_11648.pdf [accessed 2019-01-20]
7. Hsu WC, Chiang CH, Yang SC. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
8. Chiang CH, Yang SC, Hsu WC. Development and validation of the e-health literacy scale and investigation of the relationships between e-health literacy and healthy behaviour among undergraduate students in Taiwan. Formosa J Mental Health 2015;28(3):389-420.
9. Beşer A, Bahar Z, Büyükkaya D. Health promoting behaviors and factors related to lifestyle among Turkish workers and occupational health nurses' responsibilities in their health promoting activities. Ind Health 2007 Jan;45(1):151-159 [FREE Full text] [Medline: 17284887]
10. Sanlier N, Arslan S, Buyukgenc N, Toka O. Are eating behaviors related with by body mass index, gender and age? Ecol Food Nutr 2018;57(4):372-387. [doi: 10.1080/03670244.2018.1493470] [Medline: 29979919]
11. Christian H, Giles-Corti B, Knuiman M, Timperio A, Foster S. The influence of the built environment, social environment and health behaviors on body mass index. Results from RESIDE. Prev Med 2011;53(1-2):57-60. [doi: 10.1016/j.ypmed.2011.05.004] [Medline: 21609730]
12. Mellin AE, Neumark-Sztainer D, Story M, Ireland M, Resnick MD. Unhealthy behaviors and psychosocial difficulties among overweight adolescents: the potential impact of familial factors. J Adolesc Health 2002 Aug;31(2):145-153. [doi: 10.1016/s1054-139x(01)00396-2] [Medline: 12127384]
13. Monneuse MO, Bellisle F, Koppert G. Eating habits, food and health related attitudes and beliefs reported by French students. Eur J Clin Nutr 1997 Jan;51(1):46-53 [FREE Full text] [doi: 10.1038/sj.ejcn.1600361] [Medline: 9023467]

14. Gutiérrez-Pliego LE, Camarillo-Romero EDS, Montenegro-Morales LP, Garduño-García JDJ. Dietary patterns associated with body mass index (BMI) and lifestyle in Mexican adolescents. BMC Public Health 2016 Aug;16(1):850 [FREE Full text] [doi: 10.1186/s12889-016-3527-6] [Medline: 27549220]
15. Amruth M, Kumar A. A cross-sectional study on BMI and eating habits among students in a medical college in Kerala. Int J Community Med Public Health 2019 Feb;6(3):1285-1294. [doi: 10.18203/2394-6040.ijcmph20190627]
16. Howarth NC, Huang TTK, Roberts SB, Lin BH, McCrory MA. Eating patterns and dietary composition in relation to BMI in younger and older adults. Int J Obes (Lond) 2007 Apr;31(4):675-684. [doi: 10.1038/sj.ijo.0803456] [Medline: 16953255]
17. Kamtsios S, Digelidis N. Physical activity levels, exercise attitudes, self-perceptions and BMI type of 11 to 12-year-old children. J Child Health Care 2008 Sep;12(3):232-240. [doi: 10.1177/1367493508092510] [Medline: 18678585]
18. Taheri S, Lin L, Austin D, Young T, Mignot E. Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. PLoS Med 2004 Dec;1(3):e62 [FREE Full text] [doi: 10.1371/journal.pmed.0010062] [Medline: 15602591]
19. Hung HC, Yang YC, Ou HY, Wu JS, Lu FH, Chang CJ. The association between self-reported sleep quality and overweight in a Chinese population. Obesity (Silver Spring) 2013 Mar;21(3):486-492 [FREE Full text] [doi: 10.1002/oby.20259] [Medline: 23592657]
20. Chen YT, Lu TC, Cao WY, Huang SC, Chen YH. Exercise habit and sleep quality of Asia-Pacific Institute of Creativity students. Journal of Sport, Leisure and Hospitality Research 2017;12(1):1-14.
21. Peltzer K. Health behaviour among black and white South Africans. J R Soc Promot Health 2002 Sep;122(3):187-193. [doi: 10.1177/146642400212200316] [Medline: 12391834]
22. Yang SC, Luo YF, Chiang CH. Electronic health literacy and dietary behaviors in Taiwanese college students: cross-sectional study. J Med Internet Res 2019 Nov;21(11):e13140 [FREE Full text] [doi: 10.2196/13140] [Medline: 31769760]
23. Greene GW, Schembre SM, White AA, Hoerr SL, Lohse B, Shoff S, et al. Identifying clusters of college students at elevated health risk based on eating and exercise behaviors and psychosocial determinants of body weight. J Am Diet Assoc 2011 Mar;111(3):394-400. [doi: 10.1016/j.jada.2010.11.011] [Medline: 21338738]
24. Taiwan Ministry of Health and Welfare. Body Mass Index Test. 2018. URL: https://health99.hpa.gov.tw/OnlinkHealth/Onlink_BMI.aspx [accessed 2019-01-12]
25. Chen MC, Yaung CL. Determinates of sleep quality and its correlation with students in a university. Chinese Journal of School Health 2008;53:35-55.
26. Lund HG, Reider BD, Whiting AB, Prichard JR. Sleep patterns and predictors of disturbed sleep in a large population of college students. J Adolesc Health 2010 Mar;46(2):124-132. [doi: 10.1016/j.jadohealth.2009.06.016] [Medline: 20113918]
27. Yang YC, Wu T, Wang YC, Chuang CH, Su YL, Wang PC, et al. Investigation and analysis of health behaviors among residents attending community health promotion activities in Da-an District, Taipei City. Taiwan Family Medicine Research 2009;7(2):54-69.

Edited by G Eysenbach; submitted 30.12.19; peer-reviewed by CF Huang, S Khan, RS DeJesus; comments to author 10.02.20; revised version received 17.03.20; accepted 26.03.20; published 11.06.20.

Please cite as:
Hsu WC, Chiang CH
Effect of BMI and Perceived Importance of Health on the Health Behavior of College Students: Cross-Sectional Study
J Med Internet Res 2020;22(6):e17640
URL: https://www.jmir.org/2020/6/e17640
doi:10.2196/17640
PMID:32525487

Original Paper

Developing and Applying a Formative Evaluation Framework for Health Information Technology Implementations: Qualitative Investigation

Kathrin Cresswell[1*], PhD; Robin Williams[2*], PhD; Aziz Sheikh[1*], OBE, MD

[1]Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

[2]Institute for the Study of Science, Technology and Innovation, The University of Edinburgh, Edinburgh, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Kathrin Cresswell, PhD
Usher Institute
The University of Edinburgh
Teviot Place
Edinburgh,
United Kingdom
Phone: 44 1316508102
Email: kathrin.beyer@ed.ac.uk

Abstract

Background: There is currently a lack of comprehensive, intuitive, and usable formative evaluation frameworks for health information technology (HIT) implementations. We therefore sought to develop and apply such a framework. This study describes the Technology, People, Organizations, and Macroenvironmental factors (TPOM) framework we developed.

Objective: The aim was to develop and apply a formative evaluation framework for HIT implementations, highlighting interrelationships between identified dimensions and offering guidance for implementers.

Methods: We drew on an initial prototype framework developed as part of a literature review exploring factors for the effective implementation of HIT. In addition, we used qualitative data from three national formative evaluations of different HIT interventions (electronic health record, electronic prescribing, and clinical decision support functionality). The combined data set comprised 19 case studies of primarily hospital settings, and included 703 semistructured interviews, 663 hours of observations, and 864 documents gathered from a range of care settings across National Health Service (NHS) England and NHS Scotland. Data analysis took place over a period of 10 years and was guided by a framework informed by the existing evidence base.

Results: TPOM dimensions are intimately related and each include a number of subthemes that evaluators need to consider. Although technological functionalities are crucial in getting an initiative off the ground, system design needs to be cognizant of the accompanying social and organizational transformations required to ensure that technologies deliver the desired value for a variety of stakeholders. Wider structural changes, characterized by shifting policy landscapes and markets, influence technologies and the ways they are used by organizations and staff.

Conclusions: The TPOM framework supports formative evaluations of HIT implementation and digitally enabled transformation efforts. There is now a need for prospective application of the TPOM framework to determine its value.

(J Med Internet Res 2020;22(6):e15068) doi:10.2196/15068

KEYWORDS

health information technology; evaluation; sociotechnical

Introduction

Health systems worldwide are prioritizing the implementation of health information technology (HIT) in the quest to address some of health care's greatest challenges, including aging populations living with long-term conditions, persistent variations in the quality of care, and rising health care costs [1,2]. Although there is general agreement that HIT has the potential to improve safety, quality, and efficiency [3], large-scale HIT implementations require significant upfront investment, benefits are likely to materialize slowly, and those who put in most of the effort are often not those who benefit directly [4].

Such social and organizational challenges, which vary across contexts and technological functionalities, are often hard to navigate and predict for those managing change [5,6]. Despite efforts to identify success factors to guide implementation efforts, there is no simple recipe for success [7].

Formative evaluations can help to navigate these challenges. They can assist decision makers in moving from reactive to proactive strategies and identifying appropriate metrics to establish baselines and measure progress. In addition, they can help decision makers learn lessons more rapidly within the time frame of a project life cycle [8-10]. Formative evaluations can identify emerging unintended consequences and thereby, for instance, help to avert potential adverse outcomes for patient safety (eg, those arising from shortcomings in design, implementation strategy, and work practices) [11,12]. Evaluators should ideally work closely with strategic decision makers to keep projects on track and identify potential risks and mitigation strategies as early as possible.

However, despite these potential benefits, there is limited expertise in health services in conducting such formative evaluations. Existing organizational implementation evaluations, if conducted at all, often take place after an implementation has gone wrong, and use suboptimal methodologies.

Evidence-based frameworks to guide organizations in conducting HIT implementation evaluations have the potential to be helpful. A number of health informatics scholars have recently recognized this gap and developed various frameworks, some of which are summarized in Table 1 [13-19]. Some factors, such as user engagement and leadership, are well established in the change management literature. Other factors, such as political and market dimensions, have more recently received increasing recognition in shaping HIT implementations. The proliferation of frameworks poses a challenge for implementers seeking to navigate the literature and this paper seeks to integrate these frameworks.

Table 1. Examples of existing health information technology evaluation frameworks.

Framework	Key characteristics	Reference
Nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework	This framework includes the following domains: the condition or illness, the technology, the value proposition, the adopter system, the organization(s), the wider context, and the interaction and mutual adaptation between all these domains over time.	Greenhalgh et al [13]
Framework for Evaluation of Informatics Tools	This framework includes the following stages: specification and needs requirements, component development, integration of system into a clinical setting, and routine use of a system.	Kaufman et al [14]
Health Information Technology Evaluation Toolkit	This framework includes the following dimensions: articulating goals of the project, understanding stakeholders, and benefits measurement.	Cusack and Poon [15]
Health Information Systems: human, organization, and technology-fit factors (HOT-fit)	This framework focuses on the fit between technological, human, and organizational dimensions.	Yusof et al [16]
Health Information Technology Reference-based Evaluation Framework (HITREF)	This framework includes 6 dimensions: structural quality, functional quality, effects on quality processes, effects on outcome quality of care, unintended consequences, and barriers and facilitators.	Sockolow et al [18]

Based on over 10 years of experience, we set out to update current thinking about formative evaluation frameworks. Drawing on the existing literature, our aim was to develop and apply a formative evaluation framework for HIT implementations that would offer guidance for implementers. In this study, we will present our experiential conclusions and highlight interrelationships between identified dimensions.

Methods

Description of the Data Set

We have led a series of qualitative, theoretically informed case studies of different HIT implementations in the context of national formative evaluations. These included electronic health records (EHRs), clinical decision support (CDS) systems, and a combination of CDS and computerized physician order entry (CPOE) systems [4,20,21]. Our ongoing involvement as principal investigators and researchers in these various formative evaluations provided a platform for understanding HIT implementation challenges.

Table 2 shows our data set, consisting of qualitative data collected between 2009 and 2018 by our research teams that included 11 social scientists. AS was the principal investigator on two of these projects [4,20], KC was the principal investigator on one [21], and RW was a senior adviser on all three [4,20,21]. We have published several papers from these evaluations, including both primary research and sets of evaluation recommendations based on the literature [4,20,21].

KC collected some primary data (approximately 100 interviews and 60 hours of observations on [20], 40 interviews on [4], and 14 interviews on [21]). The majority of case study sites (18 of 19) included hospital settings implementing EHR and CDS/CPOE functionality in the English National Health Service (NHS).

Table 2. Data set informing the development of the evaluation framework.

Project	Data set	Timeline
National evaluation of the implementation of electronic health records in secondary care in England	12 longitudinal qualitative case studies: 431 interviews, 590 hours of observations, 234 sets of field notes, and 809 documents	February 2009 to January 2011
National evaluation of the implementation of clinical decision support/computerized physician order entry systems in English hospitals	6 longitudinal qualitative case studies: 242 interviews, 32.5 hours of observations, and 55 documents	December 2011 to March 2016
National evaluation of a pilot decision support platform in Scottish primary care	30 interviews and 8 nonparticipant ethnographic observations	May 2018 to October 2018

Sampling Overview

We defined a case as an organization implementing relevant functionality within the boundaries of an organizational setting. We sampled hospitals for maximum variation in relation to geographical location, size, implementation strategy, technological systems, and governance structures [22].

We sampled individual participants through a combination of convenience and snowball approaches with key local gatekeepers facilitating initial contacts [23]. Participants in case study settings included representatives with varying degrees of seniority from a range of clinical professions (medical, nursing, pharmacy, and allied health care professionals), managerial, and IT support staff. We also collected data from relevant policymakers and system vendors, in order to gain insights into the wider market and policy dynamics in which local implementations took place.

Data Collection Overview

We collected data between February 2009 and October 2018. The majority of data consisted of digitally audio-recorded semistructured qualitative interviews (mainly face-to-face, some by telephone). These interviews explored expectations and experiences of implementing, using, and developing the new technology (depending on the background of the interviewee). Although interview guides varied with the specific focus functionality examined, key issues explored included the following: current systems, strategies, and organizational setup; views on potential system benefits and barriers to achieving these; and future directions and visions.

In many cases (18 of 19 case studies), we sought to interview participants longitudinally (ie, before the implementation of the system, during the implementation, and once they had time to get used to the HIT system).

Observations were nonparticipant in nature, opportunistic, and involved attending relevant strategic meetings (where the researcher took notes) or following a particular activity (eg, doctors using a specific system). Observations explored technological deployments in real-world contexts.

Documents consisted of minutes of strategic meetings, summaries of lessons learned, and business cases. These provided insights into planned activities and local narratives surrounding implementation.

Development of the Framework

We began by conducting a review of existing frameworks and undertook a systematic literature review to explore which factors are important for the effective implementation of HIT [24]. The resulting prototype coding framework was iteratively refined over time and throughout projects. It included a number of dimensions and factors that formed the basis for coding qualitative data collected throughout case studies. In this process, we also allowed additional categories to emerge inductively [25]. Case studies were initially coded separately and then integrated iteratively across functionality (EHRs, CDS/CPOE in hospitals or what is known as ePrescribing in the United Kingdom, and CDS). This resulted in development of the prototype coding framework into a more comprehensive evaluation framework, which was synthesized to reflect the most pertinent categories and updated in light of the current literature. Here, our focus was on breadth rather than depth, aiming to produce a comprehensive overview of various stakeholder perspectives. We gave particular attention to stakeholders who were underrepresented (eg, vendors and administrative staff).

Our analysis and development of the novel framework was informed theoretically by the sociotechnical approach, structuration theory, the social shaping of technology, and the theory of the diffusion of innovations [26-29]. The final framework was agreed upon through iterative discussion.

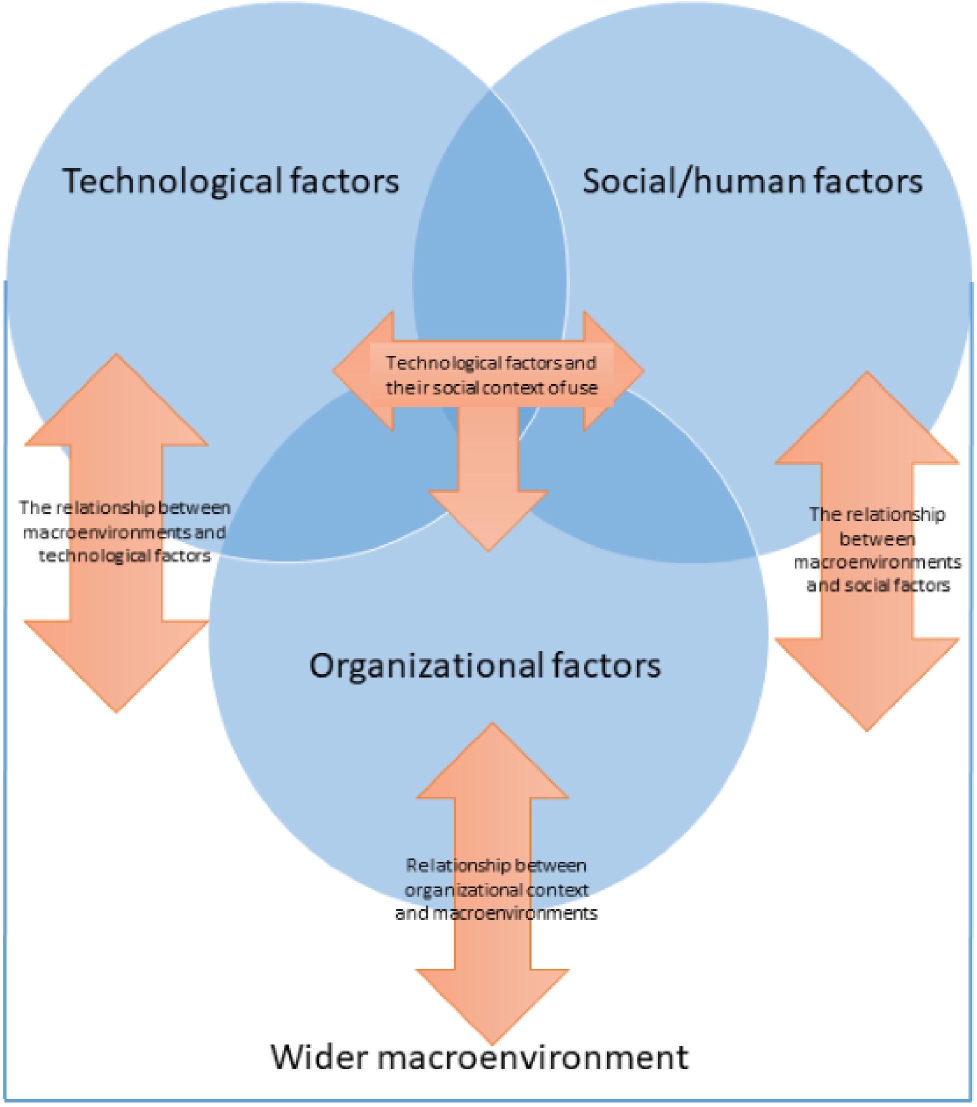
Results

We observed some commonalities across diverse settings and technological functions. The evaluation framework that has emerged from this work (Table 3, Figure 1) tackles important characteristics of the implementation landscape, where a range of technological, people (social/human), organizational, and wider macroenvironmental factors play an important role. Table 3 illustrates the key considerations in each of the dimensions. Figure 1 shows the interrelationship between the dimensions (technology, people, organizational, and macroenvironmental) and the various subcategories within each of these that need to be considered when implementing HIT.

Table 3. The Technology, People, Organizations, and Macroenvironmental factors (TPOM) framework, with example descriptions of dimensions.

Factor and dimension	Description
Technological factors	
Usability	What is the ease of use and learnability of the technology?
Performance	Does the technology function as intended by developers?
Adaptability and flexibility	Can system design be changed to suit emerging needs?
Dependability	Is the system reliable and stable?
Data availability, integrity, and confidentiality	Is data in the system available, accessible, and usable for those who need it?
Data accuracy	Is the data in the system accurate?
Sustainability	Is use of the technology sustainable?
Security	Is the system secure?
Social/human factors	
User satisfaction	Who are the users? Are users satisfied with the technology?
Complete/correct use	Are features and functionality implemented and used as intended?
Attitudes and expectations	What benefits do users expect from using the technology and how can these be measured?
Engagement	Are users actively engaged in implementation, adoption, and optimization?
Experiences	Do users have negative experience with previous technologies?
Workload/benefits	Are the benefits and efforts relatively equal for all stakeholders?
Work processes	Does the system change relationships with patients, patterns of communication, and professional responsibilities (eg, increase of administrative tasks)?
User input in design	Is there effective communication between designers, information technology staff, and end users, as well as between management and end users?
Organizational context	
Leadership and management	Are management structures to support the implementation adequate?
Communication	Are aims, timelines, and strategy communicated?
Timelines	Are implementation timelines adequate?
Vision	What benefits do organizations expect from implementing the technology and how can these be measured? Is a coherent and realistic vision driving developments?
Training and support	Is the training adequate and realistic?
Champions	Are champions and boundary spanners utilized?
Resources	Is implementation adequately resourced? (includes technology, change management, and maintenance)
Monitoring and optimization	Is system performance and use monitored and optimized over time? Are lessons learned captured and incorporated in future efforts?
Wider macroenvironment	
Media	How is the technology viewed by the media and by the public? How does the organization view/manage media relations?
Professional groups	How is the technology viewed by professional groups?
Political context	What benefits do policymakers expect from the technology and how can these be measured? What is the national approach to achieving interoperability and does the system align with this? Is there a coherent vision, consistent approach, and a clear direction of travel, allowing a degree of local input?
Economic considerations and incentives	Are there clear incentives for organizations and users to implement? (eg, improvements in quality of care) Is sufficient funding in place to support the initiative?
Legal and regulatory aspects	Have legal and regulatory frameworks been established?
Vendors	Is vendor management effectively organized?
Measuring impact	Are various stakeholders working together to define, validate, test, and refine outcome measures and measurement strategies? Are outcome measures important, clinically acceptable, transparent, feasible, and usable?

Figure 1. Diagram illustrating the evaluation framework.



None of these dimensions outranks the others. The relationship between dimensions influences how implementation, adoption, optimization, and maintenance processes unfold over time. Below, we illustrate some examples of the interrelationship between the different overarching TPOM dimensions (technology, people, organizational, and macroenvironment; Figure 1). The dimensions identified are not intended to provide detailed support for the different components of the framework, but rather to illustrate the relationships and interdependencies of the framework components.

Technological Factors and their Social Context of Use

We observed that, irrespective of the technology, systems needed to be usable, stable, and reliable (dependability), hold data securely (security), and only allow those with appropriate access rights to view confidential data (confidentiality). If systems and data within them were not dependable, users and implementers tended to lose confidence in their system choice.

> *…we are now questioning whether [name] is the right solution for our high secure service…* [Manager, EHR Evaluation]

However, it was also apparent that the design of technologies never occurred in isolation of the social and organizational context of use. This was exemplified through the many different ways in which technologies transformed how users worked in often unanticipated ways (eg, by making data entry more cumbersome; see the Social/human dimension, which includes “User satisfaction,” “Workload/benefits,” and “Work processes,” in Table 3), but also by how different technologies were optimized to suit organizational and user requirements over time (“Adaptability” in Table 3).

> *We had customized the system over a significant period of time to make it usable...* [Clinical Lead, EHR Evaluation]

System usability was the most important prerequisite for successful adoption. If users had to navigate a large number of interfaces and had difficulty finding relevant data, and if there was a general lack of intuitiveness, this slowed down their work (“Workload” and “Work processes” in Table 3).

> *All our doctors and nurses are having to work harder now, because we are having to see the same number of patients with less time, because you are spending*

> *more time on a computer now.* [Consultant, EHR Evaluation]

As a result, to not disrupt everyday delivery of care, staff had to employ workarounds ("Complete/correct use" in Table 3).

> *The staff just create workarounds of the system and some of them are ingenious and have gone on to change the system in a good way, but some aren't as helpful.* [Pharmacist, CDS/CPOE Evaluation]

Although changes to system design could address these issues, modifications were in many instances cumbersome and lengthy ("Adaptability and flexibility" in Table 3). Vendors, in turn, struggled to find a balance between tailoring of applications to local needs (achieved through local configuration) and developing commercially viable generic system versions ("Vendors" in Table 3).

> *So when [vendor] deliver the product it will have a form designer so you will be able to go in the back end […]so the focus is to, for most of it make sure that it's done through configuration not through software changes.* [Manager, CDS/CPOE Evaluation]

The Relationship Between Organizational Context and Macroenvironments

Organizations tended to employ two distinct strategies to approach system implementation. These involved both "top-down" and "bottom-up" management approaches, as well as technology strategies that involved transformative change combined with automating existing processes. "Top-down" and "bottom-up" strategies occurred at two different levels. These included the following: (1) technology design/procurement (in the English National Programme for IT (NPfIT) characterized by centralized procurement; "Political context" in the Wider macroenvironment dimension); and (2) technology implementation strategy within the organization ("Leadership and management" in the Organizational dimension).

Balancing these tensions was instrumental for the perceived success of initiatives; this is not recognized in many existing frameworks. For example, although there was a perceived need for clinicians to be heavily involved in strategy, implementation, and deployment ("Communication" and "Vision"), stakeholders also recognized that some decisions at an organizational level had to satisfy the needs of diverse stakeholders, including those outside the immediate hospital environment ("Adaptability and flexibility" in Table 3).

> *Everybody wants their own changes so some of the changes are the [management] have decided to do it that way but that doesn't mean Consultant A thinks that's the right way, so there is often not acceptance.* [Pharmacist, CDS/CPOE Evaluation]

The Relationship Between Macroenvironments and Technological Factors

It was very clear throughout our work that wider structures had a significant impact on organizational processes, ways of working, and technologies (see the Macroenvironmental dimension in Table 3). These included tensions between the long time frames needed to achieve transformation (5 to 10 years) and the episodic funding schemes (2- to 3-year programs; "Timelines" in Table 3). We also observed changing policy landscapes that involved a high turnover in senior staff and accompanying changes in visions of digital care and available funding (see "Political context" in Table 3). For example, when we began our work, England's national strategy of implementing centrally procured systems had just started [30]. However, during the CDS/CPOE evaluation, there was an increasing focus on local involvement in decision making, driven by the demise of the NPfIT and increased economic pressures resulting from the global recession. When we completed data collection, there was again a growing recognition that national guidance was crucial for promoting implementation progress and interoperability [31].

> *I've got a concern that if one of those two parties come into power and it seems highly likely that they will, that the National Programme might be closed and [system] might be shut down and what then happens…do they close the whole of the National Programme in which case, do we go back to where we were eight years ago?* [Manager, EHR Evaluation]

In addition, we observed shifts in market and vendor structures and the technologies themselves. These are not sufficiently accounted for in existing frameworks (see "Vendors" in the Wider macroenvironment dimension in Table 3). During the NPfIT, we observed a limited number of large vendors pushing other players out of the English/UK market, although the dynamics changed after the demise of NPfIT and the termination of associated contracts. This gave way to a more vibrant vendor landscape (although the market is still not very open to new vendors), which has important implications for organizations and users as they can only procure technologies from those that are currently available.

> *With the breakdown of [NPfIT] we are now seeing a lot more [hospitals] looking to take advantage of electronic prescribing and we're seeing an increased level of interest at this time to see if they can do this because effectively they've put the infrastructure in…* [Vendor, CDS/CPOE Evaluation]

We also saw how technologies were refined "in use" and through close working relationships between vendors and users over long periods ("Adaptability" in the Technology dimension in Table 3), which helped vendors/system designers overcome their limited knowledge of the social context of use. We have repeatedly seen the formation of vendor/user groups and observed how these helped to actively shape designs and markets ("User input in design").

Discussion

Summary of Findings

We have developed an evaluation framework for implementers of HIT initiatives to guide implementation and optimization of functionality (Table 3). Although this draws on formative work, it can also guide summative evaluations. The TPOM framework includes key issues to consider in relation to technological, social/human, organizational, and wider macroenvironmental dimensions. Our ongoing work has shown that these dimensions are intimately related. Technologies never exist in isolation; it is therefore critical to appreciate that technological change will be accompanied by transformations in social groups, organizations, and the wider landscapes in which these are situated, including health policy, economic climates, and the development of markets.

Strengths and Limitations

This work has drawn on a substantive composite qualitative data set collected over a long time frame. Therefore, it helped us to assess which dimensions were relatively stable over time, and only these were included in the TPOM framework. Insights presented here are views derived following careful critical reflections on a series of evaluations.

Although we acknowledge that many dimensions could be included, we deliberately attempted to keep the themes and subthemes manageable, thereby addressing a key issue in health technology evaluation: the usability of evaluation tools. We do not claim to capture all factors that play a role in implementation, adoption, and optimization of HIT, nor do we claim that our TPOM framework will provide a recipe for success. However, its pragmatic use in implementation and evaluation activity is likely to improve processes by prompting implementers to consider the most important dimensions influencing outcomes, thereby reducing unintended consequences and maximizing value. The framework now needs to be applied prospectively to confirm its utility across settings and regions. As such, we hope that it will provide a solid foundation for other countries to develop their own evaluation frameworks.

A key challenge faced by most existing evaluation frameworks, including ours, is that they neglect to account for the dynamic relationship between social and technological dimensions of change. As this relationship is a process, it does not lend itself well to presentation in 2D pragmatic evaluation tools. We have illustrated this dynamism in the Results section, drawing on concrete examples.

Integration of Findings with the Current Literature

Many empirical studies of HIT implementation are primarily concerned with evaluating impact and therefore emphasize quantitative measurements guided by benefits realization frameworks [32,33]. Recent evaluation frameworks have expanded this limited focus to include a more in-depth appreciation of the interplay of social and technological factors shaping implementations. However, although acknowledging the complexity of the process, these nuanced frameworks tend to concentrate on one particular local and situated aspect of technology implementation, thereby neglecting the role of wider structuring conditions in shaping developments [16,34-37].

Others have considered wider structuring conditions, but the tools developed lack intuitiveness, usability, and practical applicability. For example, some frameworks that are designed to shed light on sociotechnical processes can become abstract and difficult to apply by those without academic backgrounds (which arguably includes the vast majority of those implementing change in health system settings) [38]. Others have become so complex that they may include a myriad of relevant dimensions, but this attempt to capture everything may result in a level of complexity that undermines the usefulness of the framework as a tool [13,26]. The challenge is to avoid Lewis Carroll's cartographer's dilemma of needing to make a comprehensive map on the same scale as the mapped landscape, which then no longer helps users navigate [39]. Our framework has sought to address this dilemma.

There are several commonalities with existing frameworks, including a recognition of key technology, human and organizational dimensions, and their interrelationship [13-16,18]. However, TPOM is not condition- or illness-specific [13]. It is not concerned with the likelihood of the technology being adopted and its spread [13], but it is designed to help implementers of technology consider how implementation is progressing, the potential emerging risks, and what aspects therefore need attention to facilitate adoption. Implementers can apply TPOM to any HIT project at any stage of implementation. It builds on other evaluation frameworks that take into account the microcontext of use [14-16,18], to include consideration of wider macroenvironment dimensions that influence implementation and adoption. It is not concerned with management tools, but with alignment of perspectives [15]. As such, it does not provide a "recipe for success"; rather, it is a tool designed to help implementers navigate a complex landscape with many conflicting agendas and considerations. When problems or risks are identified with the help of the framework, these can be systematically targeted to facilitate implementation and adoption.

Policy Recommendations and Implications for Practice Emerging From This Work

Pragmatic formative evaluation frameworks can help to understand areas for potential improvement, benefits, and ways to streamline processes associated with technology implementation in health care settings. Evaluations need to move away from simple benefit realization approaches (that attempt to identify and measure benefits at the end of an implementation) toward formative evaluations that help the stakeholders involved adjust strategy along the way. Formative evaluations are part of a shift toward an evolutionary model where evaluation is a resource for faster and more effective learning.

Our proposed framework is a guide for implementers of technological change initiatives, to assist in planning, or during implementations of HIT initiatives in health care settings. We invite those who use it to suggest changes in both content and usability, as this will help to maximize its use and application. In due course, we hope to be able to draw on a range of data collected through the lens of the framework in different settings, refine it, and develop new insights in relation to each of the dimensions.

Large transformative policy programs aimed at facilitating technology implementation beyond hospitals are likely to require different evaluation frameworks, as their effects may be harder to trace and attribute.

Conclusions

We have drawn on a substantial body of data to develop the TPOM framework (Table 3), which stakeholders can use to monitor change processes and, if necessary, adjust the direction of HIT implementation projects. Going forward, a key challenge is likely to be the ongoing tension between attempts to capture the dynamics, processes, and interrelationships involved in technological change; the large number of these dimensions and their complexity; and the usability of evaluation tools by those delivering care, which is linked to their potential to have impact. We encourage prospective application of the TPOM framework to determine its value.

Acknowledgments

This article has drawn on a range of programs of independent research funded by the NHS Connecting for Health Evaluation Programme, The National Institute for Health Research, the Digital Health and Care Institute, and Scottish Government. The views expressed are those of the authors and not necessarily those of the funders.

We gratefully acknowledge all participants’ time and input, and the helpful comments of three anonymous reviewers who have significantly shaped this manuscript.

Conflicts of Interest

None declared.

References

1. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 04;362(5):382-385. [doi: 10.1056/nejmp0912825]
2. Jha AK, Doolan D, Grandt D, Scott T, Bates DW. The use of health information technology in seven nations. International Journal of Medical Informatics 2008 Dec;77(12):848-854. [doi: 10.1016/j.ijmedinf.2008.06.007]
3. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview. PLoS Medicine 2011 Jan 18;8(1):e1000387. [doi: 10.1371/journal.pmed.1000387]
4. Cresswell KM, Bates DW, Williams R, Morrison Z, Slee A, Coleman J, et al. Evaluation of medium-term consequences of implementing commercial computerized physician order entry and clinical decision support prescribing systems in two 'early adopter' hospitals. Journal of the American Medical Informatics Association 2014 Oct 01;21(e2):e194-e202. [doi: 10.1136/amiajnl-2013-002252]
5. May C, Mort M, Williams T, Mair F, Gask L. Health technology assessment in its local contexts: studies of telehealthcare. Social Science & Medicine 2003 Aug;57(4):697-710. [doi: 10.1016/s0277-9536(02)00419-7]
6. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review. International Journal of Medical Informatics 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007]
7. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. Journal of the American Medical Informatics Association 2013 Jun 01;20(e1):e9-e13. [doi: 10.1136/amiajnl-2013-001684]
8. Catwell L, Sheikh A. Evaluating eHealth Interventions: The Need for Continuous Systemic Evaluation. PLoS Medicine 2009 Aug 18;6(8):e1000126. [doi: 10.1371/journal.pmed.1000126]
9. Cresswell KM, Sheikh A. Undertaking sociotechnical evaluations of health information technologies. Journal of Innovation in Health Informatics 2014 Mar 18;21(2):78-83. [doi: 10.14236/jhi.v21i2.54]
10. Ammenwerth E, Rigby M. Evaluation of implementation of health IT. In: Evidence-Based Health Informatics May 20. Netherlands: IOS Press; 2016:11.
11. Cresswell KM, Mozaffar H, Lee L, Williams R, Sheikh A. Workarounds to hospital electronic prescribing systems: a qualitative study in English hospitals. BMJ Quality & Safety 2016 Apr 29;26(7):542-551. [doi: 10.1136/bmjqs-2015-005149]
12. Mozaffar H, Cresswell KM, Williams R, Bates DW, Sheikh A. Exploring the roots of unintended safety threats associated with the introduction of hospital ePrescribing systems and candidate avoidance and/or mitigation strategies: a qualitative study. BMJ Quality & Safety 2017 Feb 07;26(9):722-733. [doi: 10.1136/bmjqs-2016-005879]
13. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. Journal of Medical Internet Research 2017 Nov 01;19(11):e367. [doi: 10.2196/jmir.8775]
14. Kaufman D, Roberts WD, Merrill J, Lai T, Bakken S. Applying an Evaluation Framework for Health Information System Design, Development, and Implementation. Nursing Research 2006;55(Supplement 1):S37-S42. [doi: 10.1097/00006199-200603001-00007]
15. Cusack C, Poon E. Rockville, MD: Agency for Healthcare Research and Quality Oct. 2007. Health information technology evaluation toolkit URL: http://www.ucguide.org/pdf/ModuleIIpdf1.5.pdf [accessed 2020-05-28]

16. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). International Journal of Medical Informatics 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011]
17. Campbell M, Fitzpatrick R, Haines A, Kinmonth A, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
18. Sockolow PS, Bowles KH, Rogers ML. Health Information Technology Evaluation Framework (HITREF) Comprehensiveness as Assessed in Electronic Point-of-Care Documentation Systems Evaluations. Stud Health Technol Inform 2015;216:406-409.
19. Andargoli AE, Scheepers H, Rajendran D, Sohal A. Health information systems evaluation frameworks: A systematic review. International Journal of Medical Informatics 2017 Jan;97:195-209. [doi: 10.1016/j.ijmedinf.2016.10.008]
20. Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in. BMJ 2011 Oct 17;343(oct17 1):d6054-d6054. [doi: 10.1136/bmj.d6054]
21. Cresswell K, Callaghan M, Mozaffar H, Sheikh A. NHS Scotland's Decision Support Platform: a formative qualitative evaluation. BMJ Health Care Inform 2019 May;26(1):e100022 [FREE Full text] [doi: 10.1136/bmjhci-2019-100022] [Medline: 31160318]
22. Gentles S, Charles C, Ploeg J, McKibbon K. Sampling in qualitative research: Insights from an overview of the methods literature. The Qualitative Report 2015;20(11):1772-1789.
23. Marshall MN. Sampling for qualitative research. Family Practice 1996;13(6):522-526. [doi: 10.1093/fampra/13.6.522]
24. Cresswell K, Sheikh A. The NHS Care Record Service (NHS CRS): recommendations from the literature on successful implementation and adoption. Journal of Innovation in Health Informatics 2009 Sep 01;17(3):153-160. [doi: 10.14236/jhi.v17i3.730]
25. Fereday J, Muir-Cochrane E. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. International Journal of Qualitative Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
26. Sittig D, Singh H. A new socio-technical model for studying health information technology in complex adaptive healthcare systems. In: Cognitive Informatics for Biomedicine. Cham: Springer; 2015:59-80.
27. Stones R. Structuration theory. Basingstoke: Palgrave Macmillan; 2005.
28. Williams R, Edge D. The social shaping of technology. Research Policy 1996 Sep;25(6):865-899. [doi: 10.1016/0048-7333(96)00885-2]
29. Rogers EM. Diffusion of innovations. New York: The Free Press; 1995.
30. Currie WL, Guah MW. IT-Enabled Healthcare Delivery: The U.K. National Health Service. Information Systems Management 2006 Mar;23(2):7-22. [doi: 10.1201/1078.10580530/45925.23.2.20060301/92670.3]
31. Department of Health and Social Care. 2016. Using information technology to improve the NHS URL: https://www.gov.uk/government/publications/using-information-technology-to-improve-the-nhs [accessed 2020-05-28]
32. Shang S, Seddon P. A comprehensive framework for classifying the benefits of ERP systems. AMCIS Proceedings 2000;39:1005-1014 [FREE Full text]
33. Casey R, Wainwright D, Waring T. Benefits realisation of information technology in the National Health Service: a paradigmatic review. : Academic Conferences International Limited; 2015 Presented at: 9th European Conference on IS Management and Evaluation; 2015; UK p. 37.
34. Rippen HE, Pan EC, Russell C, Byrne CM, Swift EK. Organizational framework for health information technology. International Journal of Medical Informatics 2013 Apr;82(4):e1-e13. [doi: 10.1016/j.ijmedinf.2012.01.012]
35. Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: An evaluation of information system impact and contingency factors. International Journal of Medical Informatics 2014 Nov;83(11):779-796. [doi: 10.1016/j.ijmedinf.2014.06.011]
36. Yen P, Bakken S. Review of health information technology usability study methodologies. Journal of the American Medical Informatics Association 2011;19(3):413-422. [doi: 10.1136/amiajnl-2010-000020]
37. Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. Healthcare Quarterly 2007;10(1):112-6, 118. [Medline: 17326376]
38. Cornford T, Doukidis G, Forster D. Experience with a structure, process and outcome framework for evaluating an information system. Omega 1994 Sep;22(5):491-504. [doi: 10.1016/0305-0483(94)90030-2]
39. Carroll L. The Complete Illustrated Works. New York: Gramercy Books; 1982:727.

Abbreviations

CDS: clinical decision support
CPOE: computerized physician order entry
EHR: electronic health records
HIT: health information technology

NHS: National Health Service
NPfIT: English National Programme for IT
TPOM: Technology, People, Organizations, and Macroenvironmental factors

Edited by G Eysenbach; submitted 17.06.19; peer-reviewed by T Muto, E Broadbent, J Walker; comments to author 25.10.19; revised version received 14.11.19; accepted 19.04.20; published 10.06.20.

Please cite as:
Cresswell K, Williams R, Sheikh A
Developing and Applying a Formative Evaluation Framework for Health Information Technology Implementations: Qualitative Investigation
J Med Internet Res 2020;22(6):e15068
URL: http://www.jmir.org/2020/6/e15068/
doi:10.2196/15068
PMID:32519968

Original Paper

Patient-Centric Scheduling With the Implementation of Health Information Technology to Improve the Patient Experience and Access to Care: Retrospective Case-Control Analysis

Sukyung Chung[1,2], PhD, MPH; Meghan C Martinez[2], MPH; Dominick L Frosch[2], PhD; Veena G Jones[2,3,4], MD; Albert S Chan[3,4,5], MD, MSc

[1]Quantitative Sciences Unit, School of Medicine, Stanford University, Palo Alto, CA, United States
[2]Palo Alto Medical Foundation Research Institute, Palo Alto, CA, United States
[3]Palo Alto Medical Foundation, Palo Alto, CA, United States
[4]Sutter Health, Sacramento, CA, United States
[5]Center for Biomedical Information Research, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Sukyung Chung, PhD, MPH
Quantitative Sciences Unit
School of Medicine
Stanford University
1701 Page Mill Rd
Palo Alto, CA, 94304
United States
Phone: 1 4088363738
Email: schung1@stanford.edu

Abstract

Background: Cancellations and rescheduling of doctor's appointments are common. An automated rescheduling system has the potential to facilitate the rescheduling process so that newly opened slots are promptly filled by patients who need and can take the slot. Building on an existing online patient portal, a large health care system adopted an automated rescheduling system, Fast Pass, that sends out an earlier appointment offer to patients via email or SMS text messaging and allows patients to reschedule their appointment through the online portal.

Objective: We examined the uptake of Fast Pass at its early stage of implementation. We assessed program features and patient and visit characteristics associated with higher levels of Fast Pass utilization and the association between Fast Pass use and no-show and cancellation rates.

Methods: This study was a retrospective analysis of Fast Pass offers sent between July and December 2018. Multivariable logistic regression was used to assess the independent contribution of program, patient, and visit characteristics on the likelihood of accepting an offer. We then assessed the appointment outcome (completion, cancellation, or no-show) of Fast Pass offered appointments compared to appointments with the same patient and visit characteristics, but without an offer.

Results: Of 177,311 Fast Pass offers sent, 14,717 (8.3%) were accepted. Overall, there was a 1.3 percentage point (38%) reduction in no-show rates among Fast Pass accepted appointments compared to other appointments with matching characteristics ($P<.001$). The offers were more likely to be accepted if they were sent in the evening (versus early morning), the first (versus repeated) offer for the same appointment, for a slot 1-31 days ahead (versus same-day), for later in a day (versus before 10am), for a primary care (versus specialty) visit, sent via SMS text messaging (versus email only), for an appointment made through the online patient portal (versus via phone call or in-person), or for younger adults aged 18-49 years (versus those aged 65 years or older; all at $P<.001$). Factors negatively associated with offer acceptance were a higher number of comorbidities ($P=.02$) and visits scheduled for chronic conditions (versus acute conditions only; $P=.002$).

Conclusions: An automated rescheduling system can improve patients' access by reducing wait times for an appointment, with an added benefit of reducing no-shows by serving as a reminder of an upcoming appointment. Future modifications, such as increasing the adoption of SMS text messaging offers and targeting older adults or patients with complex conditions, may make the system more patient-centered and help promote wider utilization.

(J Med Internet Res 2020;22(6):e16451) doi:10.2196/16451

KEYWORDS

access to care; health information technology; appointment scheduling; patient-centered care

Introduction

Physicians and health care systems have long struggled with two concomitant but discordant conditions with regard to patient access to care. Many health systems simultaneously report long wait times for patient appointments and available appointment slots that are underutilized. Common contributing factors to this are last-minute cancellations and appointment no-shows by patients, leaving unfilled appointment slots in a physician's schedule. No-show rates have been variably reported to be as low as 2% to more than 50% of scheduled appointments [1-6], and represent an estimated annual cost to the health care system of $150 billion [6-8]. These schedule holes disrupt clinic workflows and reduce efficiency, reduce access for other patients who could have filled the slot, represent lost revenue for the health care systems, result in worse patient health outcomes, and ultimately cause dissatisfaction for patients and health care providers [9-13]. Typical countermeasures are often clinician-centric. For example, clinicians may maintain long wait queues to increase the probability of full schedules. Clinicians have relied on staff to manage waitlists of patients, but this approach is time- and labor-intensive, as clinic staff spend time calling patients to try to fill open slots [1,14,15].

Health information technology (HIT), particularly electronic health records (EHR), is often cited as a source of physician burnout, with clinical documentation requirements and other administrative tasks associated with EHR use in the United States being major contributing factors [16-18]. However, this literature often fails to recognize the counterbalancing beneficial aspects of HIT, such as enabling patients to access their health records electronically, schedule and reschedule appointments, and actively participate in shared decision-making, potentially resulting in improved care experiences and satisfaction [19-22]. An automated HIT system that leverages digital patient engagement to take advantage of available care options presents an opportunity to create a mutually beneficial scenario, addressing the need to improve practice efficiency and the patients' desire for more timely access.

In 2015, Sutter Health began piloting an automated appointment offer and fulfillment program called Fast Pass, a module developed by Epic Systems Corporation, and subsequently completed an enterprise-wide implementation in ambulatory primary care and specialty practices in 2018. Fast Pass allows a waitlisted patient to receive an automated message when an earlier appointment slot is available, and accept it or keep the original appointment. Fast Pass has significant potential to improve access and patient experience, in addition to promoting efficiency by filling clinic slots.

We therefore conducted a rigorous evaluation of the Fast Pass system. We evaluated the implementation of the system in 2018 after it had been rolled out across the Sutter Health system. The current paper describes the program features and identifies patient and visit factors associated with higher levels of Fast Pass utilization, as well as the association between Fast Pass implementation and no-show and cancellation rates. To our knowledge, this is the first study to systematically evaluate the implementation and effectiveness of an automated rescheduling system across ambulatory primary and specialty care settings.

Methods

Setting

Sutter Health is a large, not-for-profit health care system in Northern California serving more than 3 million people across 100 rural, suburban, and urban communities. The health care population represents high diversity in insurance coverage and race/ethnicity (eg, 17% Asian, 11% Hispanic/Latino), socio-demographics, and cultural backgrounds, mirroring the larger underlying catchment area. The health care system has a long history of utilizing HIT, with one affiliate becoming the first health system in the nation to implement Epic Systems Corporation's MyChart patient portal, My Health Online (MHO), in 2001.

Features of Fast Pass

In late 2015, Sutter implemented a pilot program, Fast Pass, to allow patients to be notified via email of an earlier appointment slot should one become available. Then, in early 2018, along with implementation of SMS text messaging reminders for appointments, Fast Pass was rolled out across the organization and patients were allowed to opt in to receiving Fast Pass notices via email, SMS text messaging, or both.

To use Fast Pass, a patient must first opt into the program via MHO (the online patient portal) and elect to receive notifications via email, SMS text messaging, or both (Figure 1). After a patient schedules an appointment, they can elect to be added to the waitlist for an earlier appointment slot. Once notified of an earlier slot, the patient must log on to MHO to respond. After several modifications during the initial pilot phase, the current Fast Pass system is scheduled to offer alternative appointment slots 9 times per day (Table 1). Two of these batches are sent early in the morning between 6 AM and 7:30 AM for same-day appointments and they expire after 30 minutes. One batch is sent midday (11:30 AM), also expires after 30 minutes, and is for appointment slots either that same day or up to 7 days ahead. The final six batches are all sent in the evening between 6 PM and 8:30 PM for appointments the next day up to 31 days ahead, and these offers all expire at 5:30 AM the following day. For each batch, an appointment slot was offered to 5 patients simultaneously; the offer expired automatically once it was taken on a first come, first served basis or when the time expired.

Figure 1. Flow of scheduling, rescheduling, and appointment outcomes with Fast Pass. MHO: My Health Online.

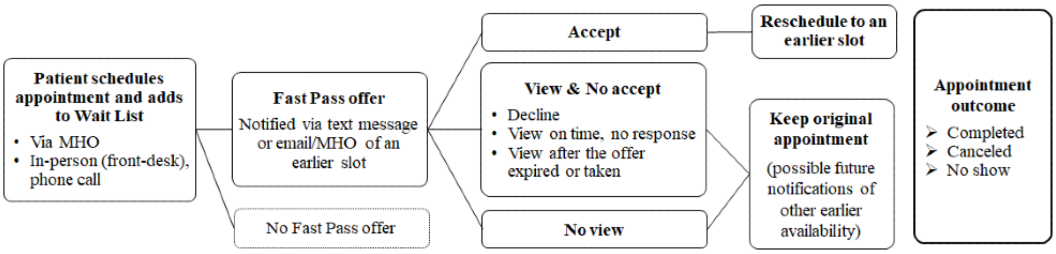


Table 1. Fast Pass offer characteristics.

	Batch 1	Batch 2	Batch 3	Batch 4	Batch 5	Batch 6	Batch 7	Batch 8	Batch 9
Time sent	6 AM	6:30 AM	11:30 AM	6 PM	6:30 PM	7 PM	7:30 PM	8 PM	8:30 PM
Time before expiration	30 minutes	30 minutes	30 minutes	11.5 hours	11 hours	10.5 hours	10 hours	9.5 hours	9 hours
Offered appointment date	Same day	Same day	Same day to 7 days ahead	1-31 days ahead	1-31 days ahead	1-31 days ahead	1-31 days ahead	1-31 days ahead	1-31 days ahead
Minimum days saved	1	1	1	3	3	3	3	3	3
Number of offers sent per slot	5	5	5	5	5	5	5	5	5

Statistical Analysis

Factors Associated With the Likelihood of Accepting a Fast Pass Offer

We used Fast Pass offers sent between July and December 2018 for this retrospective analysis. First, we examined the distribution of responses to Fast Pass offers, classified into the following: accepted, declined, viewed on time (while the offer was active) but did not respond, viewed too late (after expired or taken by another person), and did not view.

We then assessed differences in acceptance rates based on Fast Pass feature, visit type, and patient demographic and clinical characteristics. *T* test and analysis of variance (*F* test) were used to examine differences in the unadjusted proportion of accepted offers across subgroups. Multivariable logistic regression was used to assess the independent contribution of each factor on the likelihood of accepting the offer.

Fast Pass and the Likelihood of No-Shows

We assessed whether outcomes of an appointment (ie, completed, canceled, or no-show) with a Fast Pass offer (whether the appointment was rescheduled through Fast Pass or the original appointment was kept) differed from appointments without an offer. We considered a canceled appointment as happening any time prior to the scheduled appointment time. Recognizing that the impact of Fast Pass would depend on patients' responses to the offer, we analyzed two types of responses to the offer separately: "accepted" and "viewed but not accepted". For each Fast Pass offer that was accepted or viewed, we randomly selected 5 matched appointments from the pool of patients who were actively using MHO but were not offered a Fast Pass alternative for an upcoming appointment. The Fast Pass offer appointments and 5 matching no-offer appointments were exactly matched on the following characteristics: age group (0-17, 18-39, 40-54, 55-64, 65 years or older); primary care versus specialty care; scheduled appointment time (before 10 AM, 10 AM to 11:59 AM, noon to 2:59 PM, 3 PM and later); and days between appointment made and visit date (0, 1-7, 8-24, 29 days or longer). We chose the 1:5 ratio to make the best use of available data while also reducing bias, as there were very few appointments with an offer relative to those without an offer that had the same observed characteristics [23,24].

Analyses were conducted with Stata 14.2 (StataCorp). The study was reviewed and deemed to be quality improvement by the Sutter Health Institutional Review Board.

Results

Responses to Fast Pass Offers

Of 177,311 Fast Pass offers sent for 44,792 appointment slots to 38,361 patients, 8.3% (n=14,717) were accepted, 11.1% (n=19,682) were declined, 8.0% (n=14,185) were viewed on time but with no response, 53.3% (n=94,507) were viewed after the offer expired or became unavailable, and 19.4% (n=34,398) were never viewed.

Characteristics of Fast Pass Offers

Out of the 177,311 Fast Pass offers sent, most were for same-day appointments (n=48,502, 27%), appointments the next day (n=28,370, 16%), or appointments 2-7 days ahead (n=62,059, 35%), with the remainder (n=37,235, 21%) for appointments 8-31 days ahead (Table 2). If they were accepted, Fast Pass offers would have saved 8-30 days from the originally scheduled appointment in most cases (n=83,891, 47%), followed by 0-7 days (n=51,469, 29%), and 31-358 days (n=42,196, 24%). Those who accepted an offer saw their clinician on average 14.8 (SD 14.6) days sooner than initially scheduled for primary care, and 23.7 (SD 23.0) days sooner for specialty care. A quarter of offers (n=40,992, 23%) were sent during the early morning, 64,005 (36%) were sent midday, and 72,313 (41%) were sent in the evening. Fast Pass offers can be sent repeatedly for the same appointment slot, and 59,238 (33%) of appointment slots had an offer 5 or more times. See Table 2 for other patient and visit characteristics.

Table 2. Program, visit, and patient characteristics of Fast Pass offers.

Variables	Frequency, n (%)	Acceptance rate, n (%)	*P* value
Overall	177,311 (100)	14,717 (8.30)	N/A
Previous offers for the appointment			
0	43,692 (25)	7733 (17.70)	Ref
1	27,706 (16)	2798 (10.10)	<.001
2	19,971 (11)	1498 (7.50)	<.001
3	15,040 (8)	857 (5.70)	<.001
4	11,664 (7)	537 (4.60)	<.001
≥5	59,238 (33)	1244 (2.10)	<.001
Time offer was made			
6 AM to 7:20 AM	40,992 (23)	1435 (3.50)	Ref
11:30 AM to 11:55 AM	64,005 (36)	3840 (6.00)	<.001
6 PM to 8:55 PM	72,313 (41)	9473 (13.10)	<.001
Days to new appointment			
0	48,502 (27)	1795 (3.70)	Ref
1	28,370 (16)	2326 (8.20)	<.001
2-7	62,059 (35)	5709 (9.20)	<.001
8-14	14,185 (8)	1915 (13.50)	<.001
15-21	10,639 (6)	1340 (12.60)	<.001
22-31	12,412 (7)	1539 (12.40)	<.001
Days saved			
0-7	51,469 (29)	4632 (9.00)	Ref
8-30	83,891 (47)	7215 (8.60)	.01
31-358	42,196 (24)	2912 (6.90)	<.001
Scheduled via My Health Online (MHO)			
Offline	56,816 (32)	4204 (7.40)	Ref
MHO	120,495 (68)	12,411 (10.30)	<.001
Opted to receive SMS text messaging			
No SMS text messaging	122,345 (69)	9053 (7.40)	Ref
SMS text messaging	54,966 (31)	5717 (10.40)	<.001
Time of day of offered slot			
6 AM to 9:59 AM	56,740 (32)	4539 (8.00)	Ref
10 AM to 11:59 AM	44,328 (25)	3546 (8.00)	.91
Noon to 2:59 PM	46,101 (26)	3965 (8.60)	<.001
3 PM to 7 PM	31,916 (18)	2809 (8.80)	<.001
Provider specialty			
Primary care	69,833 (40)	5657 (8.10)	Ref
Specialty care	117,219 (66)	9846 (8.40)	.009
Visit type			
E/M[a], acute and chronic Dxs	23,050 (13)	2858 (12.40)	Ref
E/M, acute Dxs only	19,504 (11)	3218 (16.50)	<.001
E/M, chronic Dxs only	8866 (5)	1206 (13.60)	<.001
Preventive visit	10,639 (6)	1266 (11.90)	.23

Variables	Frequency, n (%)	Acceptance rate, n (%)	*P* value
Unknown[b]	115,252 (65)	6224 (5.40)	<.001
Charlson Comorbidity Index			
0	140,076 (79)	11,906 (8.50)	Ref
1	21,277 (12)	1787 (8.40)	.93
≥2	17,731 (10)	1188 (6.70)	<.001
Age (years)			
0-17	11,612 (7)	743 (6.40)	Ref
18-49	76,777 (43)	7447 (9.70)	<.001
50-64	45,561 (26)	3690 (8.10)	<.001
≥65	42,727 (24)	2735 (6.40)	.92
Sex			
Male	57,542 (32)	4891 (8.50)	Ref
Female	119,135 (67)	9769 (8.20)	<.001
Race/ethnicity			
Non-Hispanic white	100,414 (57)	8134 (8.10)	Ref
African American	6176 (3)	476 (7.70)	.24
Asian	30,863 (17)	2778 (9.00)	<.001
Latino/Hispanic	19,170 (11)	1591 (8.30)	.44
Other race or race unknown	20,041 (11)	1704 (8.50)	.08
Same-day or next-day slot offers	60,206 (100)	3131 (5.2)	N/A
Time offer was made			
6 AM to 7:20 AM	43,692 (68)	1433 (3.50)	Ref
11:30 AM to 11:55 AM	27,706 (13)	344 (4.40)	.001
6 PM to 8:55 PM	19,971 (20)	1385 (11.50)	<.001
Visit type			
E/M, acute and chronic Dxs	9031 (15)	768 (8.50)	Ref
E/M, acute Dxs only	7225 (12)	881 (12.20)	<.001
E/M, chronic Dxs only	3010 (5)	262 (8.70)	.69
Preventive visit	3612 (6)	275 (7.60)	.045
Unknown[b]	37,328 (62)	971 (2.60)	<.001

[a] E/M: evaluation and management coding.

[b] This category includes visits with no diagnosis falling into acute or chronic conditions and appointments later canceled or no-shows.

Bivariate Analysis of Characteristics of Accepted Fast Pass Offers

The first offer for an appointment slot was far more likely to be accepted than subsequent offers for the same original appointment (17.7% versus 2.1%-10.1% of acceptance rates, respectively; *P*<.001; Table 2). Offers sent in the evening were more than twice as likely to be accepted (13.1%) than morning offers (3.5%-6.0%; *P*<.001). There was a higher acceptance rate when the offered slot was more than one week ahead (12.4%-13.5%), 2-7 days ahead (9.2%), or the next day (8.2%), as compared to the same day (3.7%; *P*<.001). The acceptance rate was higher for offers with potentially fewer days saved (9.0% for 0-7 days versus 6.9% for 31-358 days; *P*<.001) and when the offered slot was in the afternoon (8.6%-8.8%) rather than in the morning (8.0%; *P*<.001). Acceptance rates were higher when the appointment was scheduled through MHO versus offline (10.3% versus 7.4%; *P*<.001) and among patients who opted to be contacted through SMS text messaging versus email only (10.4% versus 7.4%; *P*<.001). Compared to non-Hispanic whites, Asians were more likely to accept an offer (9.0% versus 8.1%; *P*<.001).

Patients with a high comorbidity burden were less likely to accept offers (Charlson Comorbidity Index [CCI]=0, 8.5%; CCI≥2, 6.7%; *P*<.001). Acceptance rates were highest among people aged 18-49 years (9.7%) and lower for those aged 0-17 years (6.4%; their guardian received the messages) or 50-64 years (8.1%; *P*<.001). Male patients were more likely to accept an offer than female patients (8.5% versus 8.2%; *P*<.001). Accepted visits were more likely to be for acute conditions only (16.5%) rather than for chronic conditions only (13.6%) or both acute and chronic conditions (12.4%; *P*<.001). Specialty care visits were slightly more likely to be accepted (8.4% versus 8.1%; *P*=.009), and there were no differences in acceptance rates across specialties (eg, dermatology, ob-gyn, orthopedics).

For same day or next-day offers only (N=60,206), offers sent in the evening (of the day before the opened slot) rather than in the morning (of the appointment day) were three times more likely to be accepted (11.5% versus 3.5%; *P*<.001), and patients with acute conditions only (12.2%) were more likely to accept the offer (*P*<.001) than those with both chronic and acute conditions (8.5%).

Factors Associated With Offer Acceptance in Multivariable Analysis

After controlling all other factors, findings from the adjusted, multivariable analysis (Table 3) were similar to those from the unadjusted, bivariate analysis, with a few exceptions. Preventive visit, older age (aged 65 years or older), and specialty visit were negative predictors of Fast Pass acceptance (*P*<.001). In addition, the following factors no longer had a significant association with offer acceptance in the adjusted model: potentially saved at least 31 days (versus 7 days or fewer; *P*=.33), CCI≥2 (*P*=.24), visit for chronic conditions (versus both acute and chronic conditions; *P*=.48), female (*P*=.04), aged 50-64 years (versus 0-17; *P*=.17), and Asian (versus non-Hispanic white; *P*=.08).

Table 3. Predictors of Fast Pass offer acceptance from a logistic regression (N=176,615)[a].

Variables	Odds ratio	99% CI	*P* value
Previous offers for the same appointment			
0	Ref		
1	0.55	0.52-0.59	<.001
2	0.40	0.37-0.43	<.001
3	0.30	0.27-0.33	<.001
4	0.24	0.22-0.27	<.001
≥5	0.11	0.10-0.12	<.001
Time offer was made			
6 AM to 7:20 AM	Ref		
11:30 AM to 11:55 AM	1.19	1.01-1.41	.006
6 PM to 8:55 PM	2.30	1.92-2.76	<.001
Days to new appointment			
0	Ref		
1	1.58	1.34-1.86	<.001
2-7	1.84	1.57-2.15	<.001
8-14	2.04	1.71-2.43	<.001
15-21	1.93	1.60-2.33	<.001
22-31	1.91	1.59-2.30	<.001
Days saved			
0-7	Ref		
8-30	1.06	1.00-1.12	.02
31-358	0.97	0.90-1.05	.33
Scheduled via My Health Online (MHO)			
Offline	Ref		
MHO	1.31	1.22-1.40	<.001
Opted to receive SMS text messaging			
No SMS text messaging	Ref		
SMS text messaging	1.46	1.37-1.55	<.001
Time of day of offered slot			
6 AM to 9:59 AM	Ref		
10 AM to 11:59 AM	1.12	1.05-1.20	<.001
Noon to 2:59 PM	1.16	1.09-1.24	<.001
3 PM to 7 PM	1.26	1.18-1.36	<.001
Provider specialty			
Primary care	Ref		
Specialty care	0.81	0.76-0.87	<.001
Visit type			
E/M, acute and chronic Dxs	Ref		
E/M, acute Dxs only	1.13	1.02-1.26	.002
E/M, chronic Dxs only	1.04	0.90-1.20	.48
Preventive visit	0.76	0.67-0.87	<.001
Unknown	0.28	0.25-0.30	<.001

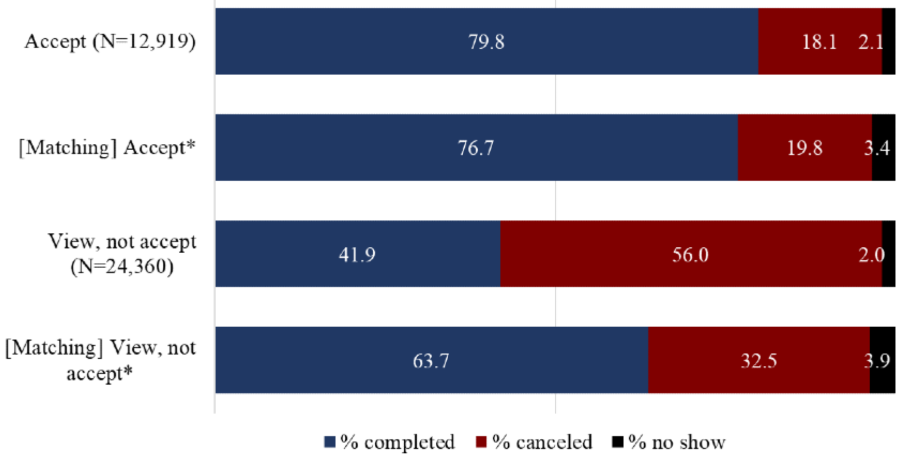
Variables	Odds ratio	99% CI	*P* value
Charlson Comorbidity Index			
0	Ref		
1	1.08	0.99-1.18	.02
≥2	0.95	0.85-1.06	.24
Age (years)			
0-17	Ref		
18-49	1.34	1.17-1.53	<.001
50-64	1.08	0.94-1.24	.17
≥65	0.81	0.69-0.93	<.001
Sex			
Male	Ref		
Female	0.95	0.90-1.01	.04
Race/ethnicity			
Non-Hispanic white	Ref		
African American	0.99	0.83-1.18	.91
Asian	0.95	0.87-1.03	.08
Hispanic/Latino	0.99	0.90-1.09	.73
Other race or race unknown	0.95	0.87-1.04	.15

[a] Indicators of clinic location were also included but not presented here.

Fast Pass and the Likelihood of Appointment Completion, Cancellation, and No-Shows

As compared to appointments with matched characteristics but without a Fast Pass offer, accepted Fast Pass appointments were more likely to be completed (79.8% versus 76.7%) and less likely to be canceled (18.1% versus 19.8%) or result in no-shows (2.1% versus 3.4%; *P*<.001; Figure 2). Additionally, Fast Pass offers viewed but not accepted were less likely to be completed (41.9% versus 63.7%) and more likely to be canceled (56.0% versus 32.5%) but less likely to be no-shows (2.0% versus 3.9%; *P*<.001).

Figure 2. Appointment outcomes by Fast Pass offer status. * indicates matching comparison groups of "accept" and "view, not accept" groups, respectively.



Discussion

Principal Findings

Fast Pass, a patient-centered application of HIT, mutually benefits patients and providers. Patients who accepted a Fast Pass offer experienced more timely access to care, with time to see a doctor reduced by 15 days in primary care and 24 days in specialty care, on average. From the provider perspective, Fast Pass offers facilitated efficient management of office visit slots by automating rescheduling processes and filling gaps in provider schedules in a timely manner.

The Fast Pass system also serves to remind the patient about an upcoming appointment, reducing no-shows by 1.3 percentage points (a 38% relative reduction or 168 appointments) and facilitating timely cancellation of appointments that are no longer needed by 1.7 percentage points (a 9% relative reduction or 219 appointments), when the offer was accepted. Even when the original appointment was kept, those who viewed the offer had almost half the no-show rate as their matched comparisons (2.0% versus 3.9%). Other studies have found that patients often miss an appointment simply because they have forgotten about it, because of long delays between scheduling an appointment and the actual visit date, or because patients have no means of cancelling [1,4,6,25,26]. Using SMS text messaging to send reminders for upcoming appointments has been shown to reduce no-show rates and increase cancellation rates, providing time for the health care system to fill that slot with another patient [1,14,15]. Though the change in no-show rates seems relatively modest, if all patients in the health care system opt into the program, it would have significant clinical implications. Applying the rate to the volume of appointments in this health care system, approximately 20 million in 2018, it translates to 260,000 no-shows that could have been avoided annually if all patients were reminded through the Fast Pass system.

Despite the potential benefits, the acceptance of Fast Pass offers was 8% of the total early appointments offered. Notably, Fast Pass offers were more likely to be accepted when the initial notification was received via SMS text messaging. The alternative to SMS text messaging of Fast Pass offers is email, which may explain the high rate of Fast Pass offers (53.3%) that were viewed after the offer expired or became unavailable. In light of protections provided by the Telephone Consumer Protection Act, which implements consumer protections from unwelcome solicitations via phone and text messaging, the organization has offered this feature to patients who have individually enrolled and consented to receiving reminders via SMS text messaging. Balancing the activation inertia of enrolling in a new program and the desire to advocate for robust consumer protections versus the desire to maximize the potential patient benefits comes with the cost of increased socio-behavioral and technical barriers to adoption. Alternative approaches not being tested here (eg, more automated enrollment for digital waitlists and SMS text messaging as well as tighter integration with conditions of registration workflows) should also be considered and evaluated in future modifications.

There are several potentially modifiable program features to improve the benefits realized from Fast Pass for those who opted into the program. Our findings suggest that offer acceptance can improve by increasing the rate of patients enrolled to receive offers delivered by SMS text messaging; sending next-day offers in the evening rather than same-day offers in the early morning; sending an offer once per appointment slot and, if not accepted, not sending another one, but instead offering it to another person; allowing more time to respond for same-day slots; and focusing on users of the online patient portal, who are more likely to view and respond to the offer. A potential modification that might help improve the acceptance rate and make the system more patient-centered is to ask patients their preference for alternative slots at the time they sign up for the waitlist. This may become feasible as more patients sign up for Fast Pass and the volume of waitlist participants is large enough to identify matching between available and preferred slots.

It is unclear, though, whether Fast Pass offers are currently utilized by those who need it most. Patients who perceive their conditions as urgent may be more likely to accept the offer, as shown with the higher acceptance rate among patients with acute conditions only and a lower rate for preventive visits. On the other hand, the acceptance rate was lower for those with multiple comorbid conditions and higher disease burdens and for older patients whose medical needs are typically greater. One potential explanation is that older adults and patients with multiple comorbidities are less likely to opt for SMS text messaging than email. To improve overall uptake of the system as well as to reduce disparities in the utilization, more outreach efforts may be necessary, especially to those who could benefit most from the system, eg, patients with higher comorbidity burdens.

Limitations

Our study has several limitations that future extended studies should address. First, though we were able to quantitatively measure the rates of acceptances and declines of Fast Pass offers, we do not understand facilitators and barriers to using the Fast Pass program from the perspective of the patient. Focus groups or patient surveys would better examine the factors that impact adoption of the program. Second, we have yet to assess whether the system reduced the workload related to rescheduling for front desk staff and schedulers, as intended. Finally, we investigated one form of an automated rescheduling system, suggesting potential modifications for improvement. A prospective study, preferably with natural experiments, should address whether such modifications would help facilitate better utilization of the system by patients who could benefit most.

Conclusions

Fast Pass, an innovative automated rescheduling system embedded in the EHR, improves patient access to care by reducing wait time for an appointment, with an added benefit of reducing no-shows and increasing timely appointment cancellation by serving as a reminder of an upcoming appointment. Future modifications, such as allowing more time to respond to the offer and targeting patients with complex conditions, may help enhance the value of the system and wider adoption of the model beyond the Sutter Health system. Identifying opportunities for the implementation of interventions such as Fast Pass can provide evidence of the full range of benefits of HIT by simultaneously improving clinician practice efficiency and patient experience and access to care.

Acknowledgments

SC had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All the analytical work and the initial manuscript was completed during her employment at Sutter Health. SC, ASC,

VGJ, and DLF, and MCM conceived of the study concept and design, and reviewed and provided critical feedback on manuscript drafts. SC and MCM led data interpretation, and framing and writing of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared. The authors are employees of the health care system where the study was conducted. The organization had no influence on the conduct or reporting of this study.

References

1. Laurent B. [Address from the president for 2010]. Rev Neurol (Paris) 2010 May;166(5):471-472. [doi: 10.1016/j.neurol.2010.05.001] [Medline: 20569791]
2. Waller J, Hodgkin P. Defaulters in general practice: who are they and what can be done about them? Fam Pract 2000 Jun;17(3):252-253. [doi: 10.1093/fampra/17.3.252] [Medline: 10846145]
3. Leong KC, Chen WS, Leong KW, Mastura I, Mimi O, Sheikh MA, et al. The use of text messaging to improve attendance in primary care: a randomized controlled trial. Fam Pract 2006 Dec;23(6):699-705 [FREE Full text] [doi: 10.1093/fampra/cml044] [Medline: 16916871]
4. Dantas LF, Fleck JL, Cyrino Oliveira FL, Hamacher S. No-shows in appointment scheduling - a systematic literature review. Health Policy 2018 Dec;122(4):412-421. [doi: 10.1016/j.healthpol.2018.02.002] [Medline: 29482948]
5. Ulmer T, Troxler C. The economic cost of missed appointments and the open access system. Gainesville, FL: University of Florida; 2006. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.513.1864 [accessed 2020-05-06]
6. Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res 2016 Jan 14;16:13 [FREE Full text] [doi: 10.1186/s12913-015-1243-z] [Medline: 26769153]
7. Toland B. No-shows cost health care system billions.: Pittsburgh Post-Gazette URL: https://www.post-gazette.com/business/businessnews/2013/02/24/No-shows-cost-health-care-system-billions/stories/201302240381 [accessed 2020-05-06]
8. Gier J. Missed appointments cost the U.S. healthcare system $150B each year.: Healthcare Innovation URL: https://www.hcinnovationgroup.com/clinical-it/article/13008175/missed-appointments-cost-the-us-healthcare-system-150b-each-year [accessed 2020-05-06]
9. Colubi MM, Pérez-Elías MJ, Elías L, Pumares M, Muriel A, Zamora AM, SEAD Study Group. Missing scheduled visits in the outpatient clinic as a marker of short-term admissions and death. HIV Clin Trials 2012;13(5):289-295. [doi: 10.1310/hct1305-289] [Medline: 23134630]
10. Husain-Gambles M, Neal RD, Dempsey O, Lawlor DA, Hodgson J. Missed appointments in primary care: questionnaire and focus group study of health professionals. Br J Gen Pract 2004 Mar;54(499):108-113 [FREE Full text] [Medline: 14965389]
11. Davies ML, Goffman RM, May JH, Monte RJ, Rodriguez KL, Tjader YC, et al. Large-Scale No-Show Patterns and Distributions for Clinic Operational Research. Healthcare (Basel) 2016 Mar 16;4(1):15-27 [FREE Full text] [doi: 10.3390/healthcare4010015] [Medline: 27417603]
12. Nguyen DL, Dejesus RS, Wieland ML. Missed appointments in resident continuity clinic: patient characteristics and health care outcomes. J Grad Med Educ 2011 Sep;3(3):350-355 [FREE Full text] [doi: 10.4300/JGME-D-10-00199.1] [Medline: 22942961]
13. George A, Rubin G. Non-attendance in general practice: a systematic review and its implications for access to primary health care. Fam Pract 2003 Apr;20(2):178-184. [doi: 10.1093/fampra/20.2.178] [Medline: 12651793]
14. Hashim MJ, Franks P, Fiscella K. Effectiveness of telephone reminders in improving rate of appointments kept at an outpatient clinic: a randomized controlled trial. J Am Board Fam Pract 2001;14(3):193-196 [FREE Full text] [Medline: 11355051]
15. Rohman L, Maruswezki D, Boyce Cam N. The impact of a text messaging service on orthopaedic clinic Did Not Attend rates. J Telemed Telecare 2015 Oct;21(7):408-413. [doi: 10.1177/1357633X15583778] [Medline: 26026176]
16. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients And Desktop Medicine. Health Aff (Millwood) 2017 Dec 01;36(4):655-662 [FREE Full text] [doi: 10.1377/hlthaff.2016.0811] [Medline: 28373331]
17. Downing NL, Bates DW, Longhurst CA. Physician Burnout in the Electronic Health Record Era. Ann Intern Med 2019 Mar 05;170(3):216-217. [doi: 10.7326/L18-0604] [Medline: 30716744]
18. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
19. Panattoni L, Chan A, Yang Y, Olson C, Tai-Seale M. Nudging physicians and patients with autopend clinical decision support to improve diabetes management. Am J Manag Care 2018 Oct;24(10):479-483 [FREE Full text] [Medline: 30325190]

20. Silvestre A, Sue VM, Allen JY. If you build it, will they come? The Kaiser Permanente model of online health care. Health Aff (Millwood) 2009;28(2):334-344 [FREE Full text] [doi: 10.1377/hlthaff.28.2.334] [Medline: 19275988]
21. Neuner J, Fedders M, Caravella M, Bradford L, Schapira M. Meaningful use and the patient portal: patient enrollment, use, and satisfaction with patient portals at a later-adopting center. Am J Med Qual 2015;30(2):105-113. [doi: 10.1177/1062860614523488] [Medline: 24563085]
22. Tai-Seale M, Downing NL, Jones VG, Milani RV, Zhao B, Clay B, et al. Technology-Enabled Consumer Engagement: Promising Practices At Four Health Care Delivery Organizations. Health Aff (Millwood) 2019 Mar;38(3):383-390. [doi: 10.1377/hlthaff.2018.05027] [Medline: 30830826]
23. Smith HL. Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies. Sociological Methodology 2016 Jun 24;27(1):325-353. [doi: 10.1111/1467-9531.271030]
24. Rubin DB. Matching to Remove Bias in Observational Studies. Biometrics 1973 Mar;29(1):159. [doi: 10.2307/2529684]
25. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]
26. Kaplan-Lewis E, Percac-Lima S. No-show to primary care appointments: why patients do not come. J Prim Care Community Health 2013 Oct;4(4):251-255. [doi: 10.1177/2150131913498513] [Medline: 24327664]

Abbreviations

CCI: Charlson Comorbidity Index
HIT: health information technology
MHO: My Health Online

Edited by G Eysenbach; submitted 30.09.19; peer-reviewed by P Zaki, D Newman, D Amiri; comments to author 26.11.19; revised version received 09.01.20; accepted 29.03.20; published 10.06.20.

Please cite as:
Chung S, Martinez MC, Frosch DL, Jones VG, Chan AS
Patient-Centric Scheduling With the Implementation of Health Information Technology to Improve the Patient Experience and Access to Care: Retrospective Case-Control Analysis
J Med Internet Res 2020;22(6):e16451
URL: https://www.jmir.org/2020/6/e16451
doi:10.2196/16451
PMID:32519970

Original Paper

Effect of Speech Recognition on Problem Solving and Recall in Consumer Digital Health Tasks: Controlled Laboratory Experiment

Jessica Chen[1], BEng, MBiomedEng; David Lyell[1], PhD; Liliana Laranjo[1], MD, MPH, PhD; Farah Magrabi[1], BEng, PhD

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia

Corresponding Author:
Jessica Chen, BEng, MBiomedEng
Centre for Health Informatics
Australian Institute of Health Innovation
Macquarie University
Level 6
75 Talavera Rd
North Ryde, NSW 2113
Australia
Phone: 61 2 9850 2400
Email: jessica.a.chen@mq.edu.au

Abstract

Background: Recent advances in natural language processing and artificial intelligence have led to widespread adoption of speech recognition technologies. In consumer health applications, speech recognition is usually applied to support interactions with conversational agents for data collection, decision support, and patient monitoring. However, little is known about the use of speech recognition in consumer health applications and few studies have evaluated the efficacy of conversational agents in the hands of consumers. In other consumer-facing tools, cognitive load has been observed to be an important factor affecting the use of speech recognition technologies in tasks involving problem solving and recall. Users find it more difficult to think and speak at the same time when compared to typing, pointing, and clicking. However, the effects of speech recognition on cognitive load when performing health tasks has not yet been explored.

Objective: The aim of this study was to evaluate the use of speech recognition for documentation in consumer digital health tasks involving problem solving and recall.

Methods: Fifty university staff and students were recruited to undertake four documentation tasks with a simulated conversational agent in a computer laboratory. The tasks varied in complexity determined by the amount of problem solving and recall required (simple and complex) and the input modality (speech recognition vs keyboard and mouse). Cognitive load, task completion time, error rate, and usability were measured.

Results: Compared to using a keyboard and mouse, speech recognition significantly increased the cognitive load for complex tasks ($Z=-4.08$, $P<.001$) and simple tasks ($Z=-2.24$, $P=.03$). Complex tasks took significantly longer to complete ($Z=-2.52$, $P=.01$) and speech recognition was found to be overall less usable than a keyboard and mouse ($Z=-3.30$, $P=.001$). However, there was no effect on errors.

Conclusions: Use of a keyboard and mouse was preferable to speech recognition for complex tasks involving problem solving and recall. Further studies using a broader variety of consumer digital health tasks of varying complexity are needed to investigate the contexts in which use of speech recognition is most appropriate. The effects of cognitive load on task performance and its significance also need to be investigated.

(J Med Internet Res 2020;22(6):e14827) doi:10.2196/14827

KEYWORDS

speech recognition software; consumer health informatics; ergonomics

Introduction

Recent advances in natural language processing and artificial intelligence have led to improvements in and widespread adoption of speech recognition technologies [1]. Speech recognition is an input modality that translates human speech into computerized text [2]. In consumer applications, speech recognition is usually applied as a way to interact with conversational agents, which are systems that mimic human conversation using text or spoken language [3,4]. Consumer conversational agents (such as Amazon Alexa and Google Assistant) can improve patient workflow by allowing patients to call nurses [5]. In health care, conversational agents have been utilized for a variety of purposes, including data collection, decision support, and patient monitoring [5-8].

Problems with the use of digital health technology represent a well-documented safety concern in the literature [9,10]. However, little is known about the problems associated with conversational agents that pose actual or potential risks of harm to consumers [9]. One study in which the participants were asked to interact with conversational agents identified significant safety concerns arising from the quality of information provided in response to health-related questions [11]. For example, incorrect information provided by a conversational agent in response to a question about the amount of alcohol that could be consumed while taking oxycodone could lead to severe harm, including death. However, few studies have evaluated conversational agents in the hands of consumers [3,4].

In other consumer-facing tools, cognitive load has been observed to be an important factor affecting the use of speech recognition technologies. Users reported finding it more difficult to think and speak at the same time when compared to typing, pointing, and clicking [12,13]. However, the effects of speech recognition use on cognitive load when performing health tasks has not yet been explored. Cognitive load is the amount of workload imposed on the brain's working or short-term memory, which has limited capacity [14,15] and a short duration [16], particularly when performing tasks requiring problem solving and recall [14,17]. An example of a problem-solving task is using basic arithmetic to calculate nutritional information, whereas recall involves memorizing and reporting exercise information. Due to the cognitive load, certain tasks may be more difficult to perform because speaking shares the same cognitive resources in the brain as those required for problem solving and recall (ie, working memory) [12]. Therefore, a possible challenge with the use of speech recognition is that it can increase the cognitive load in tasks requiring more problem solving and recall.

Although studies in other domains have investigated the effects of speech recognition use on cognitive load, to our knowledge, no study has measured its effects in digital health tasks [12,18,19]. Thus, the aim of the present study was to evaluate the use of speech recognition for documentation in consumer digital health tasks such as recording diet and exercise information in comparison to using a conventional keyboard and mouse. The following hypotheses were tested: (1) cognitive load is higher for speech recognition compared to keyboard and mouse use in complex tasks requiring more problem solving and recall; (2) the percentage of errors is higher for speech recognition compared with keyboard and mouse use in complex tasks requiring more problem solving and recall; (3) task completion time is lower for speech recognition compared with keyboard and mouse use in simple tasks requiring less problem solving and recall; and (4) speech recognition is less usable than a keyboard and mouse for both simple and complex tasks requiring more problem solving and recall.

These findings will shed light on the characteristics of consumer digital health tasks that make them most suitable for using speech recognition as an input modality.

Methods

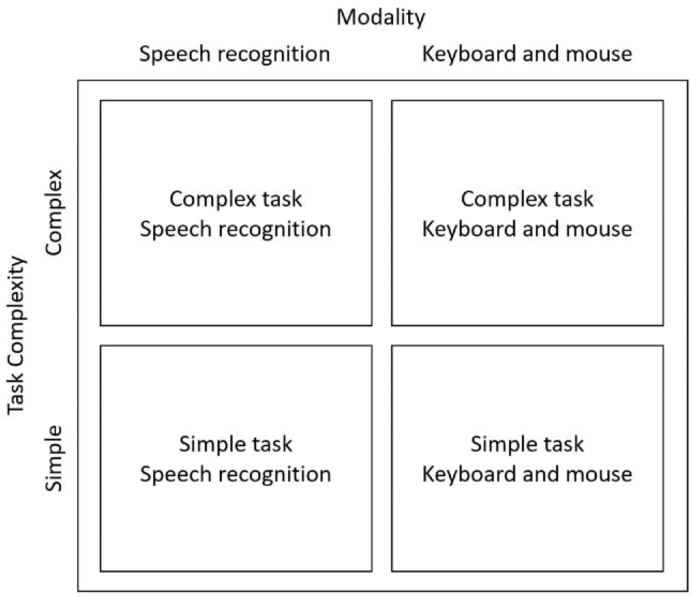
Participants

Fifty-two university students and staff participated in this study (see Results for a summary of the participant demographics). The participants were either students or staff who met the minimum English language proficiency for admission to a university program or workplace. These participants also had working knowledge of computer technology and systems as required for their degree or profession. Hence, there were no requirements regarding level of English language, health literacy, or technology proficiency for inclusion. The participants responded to advertisements sent by email or published in a university newsletter. Consenting adults aged 18 years or older were eligible to participate. Ethical approval was granted by Macquarie University's Human Research Ethics Committee to recruit people within the Faculty of Medicine and Health Sciences. Participants were not offered any remuneration or gifts to incentivize participation.

Experimental Design and Tasks

The study included two within-subject factors: human-computer interaction modality (speech recognition vs keyboard and mouse) and task complexity (simple vs complex) providing four experimental conditions (Figure 1).

Figure 1. Experimental design.



Each participant was asked to complete four consumer digital health tasks to document nutritional and exercise information: two simple and two complex tasks, using speech recognition and a keyboard and mouse to interact with a simulated conversational agent (Figure 1). For each documentation task, participants were asked to adopt a persona within a hypothetical scenario focusing on physical activity and diet, and to answer the conversational agent's questions (see Multimedia Appendix 1). The tasks were designed to assess problem solving and recall using the information provided within the scenario.

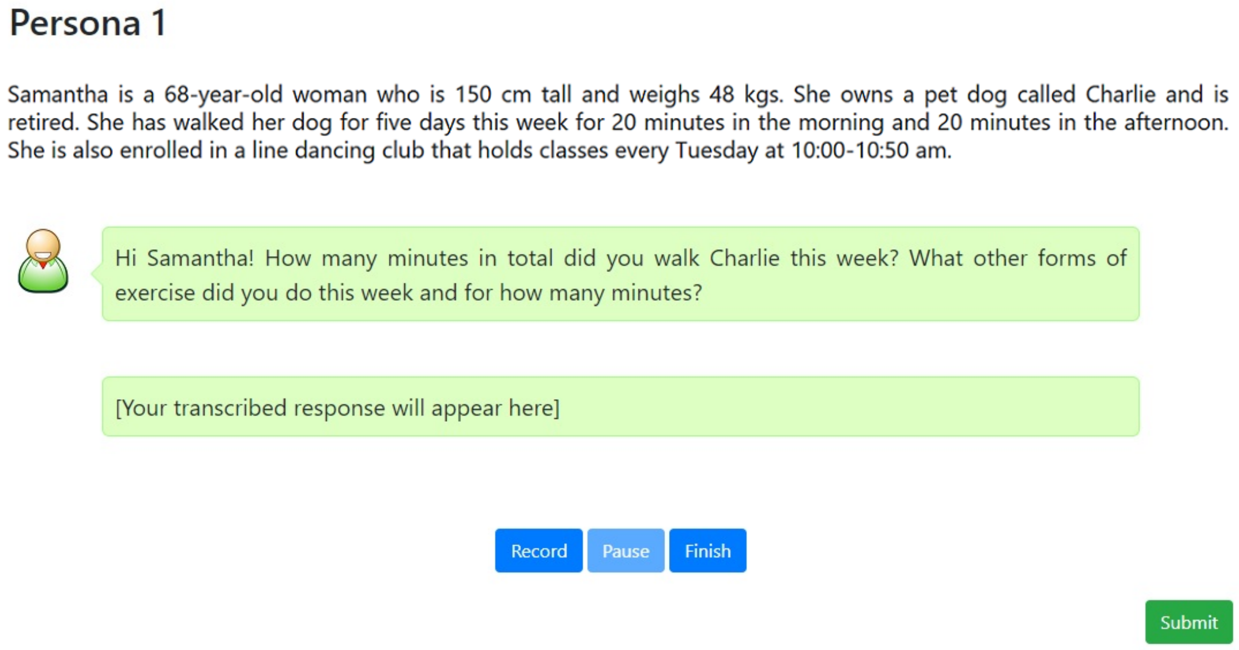
Complexity was measured by the number of information items that participants needed to manipulate in working memory. The relationship between the number of information items requiring manipulation in working memory and cognitive load has been well established in the literature [20]; that is, human performance is affected when the cognitive load is too high or exceeds the limits of working memory [21,22]. For simple tasks, the conversational agent displayed the hypothetical scenario on the same screen as the questions. Complex tasks were designed to impose a higher cognitive load by increasing the number of pieces of information needed to perform arithmetic and commit to memory (6 total items for simple tasks vs 17 total items for complex tasks; see Multimedia Appendix 1). Simple tasks required participants to problem solve using 5 items of information, including 5 days, 20 minutes, line dancing, 10:00-10.50 am, whereas complex tasks required participants to recall the same 5 items and were provided with 12 items for problem solving. A higher cognitive load was also imposed by displaying the scenario in a pop-up modal window, requiring participants to commit key information in the scenario to working memory when transferring between the windows [23]. Participants could not copy and paste their answers.

The tasks were developed in consultation with a health informatics researcher (DL) and a primary care physician (LL). Pilot testing was performed by asking 7 individuals to complete the tasks using a prototype of the system and provide feedback. Any issues were iteratively fixed before the next pilot test. This pilot testing ensured that the prescribed tasks and the system were understandable and functional. Individuals who assisted with pilot testing were excluded from participating in the experiment. The correct answers to tasks were predetermined and validated by the health informatics researcher.

Simulated Conversational Agent

Participants attended a computer laboratory at the university where a workstation was set up with the tasks running on a web app connected to a keyboard, mouse, and microphone. Presented as a text message conversation, user responses could either be typed using the keyboard and mouse or transcribed using speech recognition depending on the experimental condition (Figure 2). Participants were able to view their responses before submitting. Web Speech API, a JavaScript-based general purpose speech recognition application programming interface, was used to implement a live audio transcription function for this system [24]. To activate speech-to-text using speech recognition, the participants were instructed to press the "Record" button. The system was launched using Google Chrome on a local machine.

Figure 2. Graphical user interface of the simulated conversational agent on Google Chrome.



Persona 1

Samantha is a 68-year-old woman who is 150 cm tall and weighs 48 kgs. She owns a pet dog called Charlie and is retired. She has walked her dog for five days this week for 20 minutes in the morning and 20 minutes in the afternoon. She is also enrolled in a line dancing club that holds classes every Tuesday at 10:00-10:50 am.

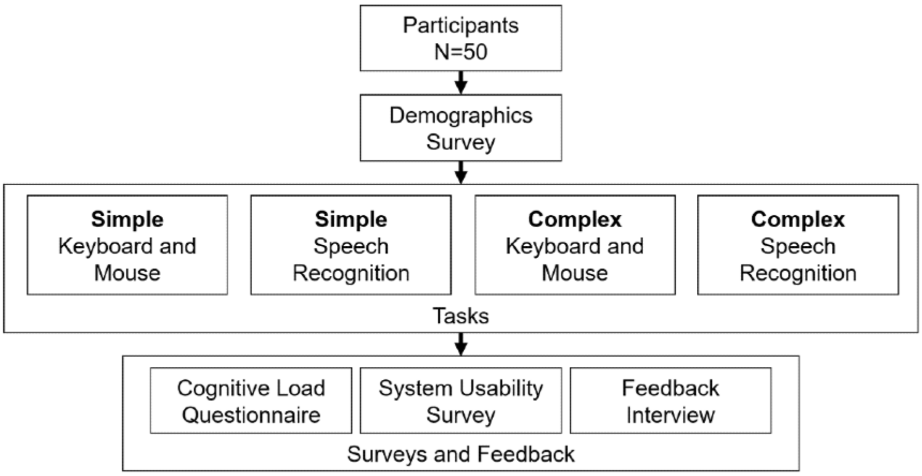
Hi Samantha! How many minutes in total did you walk Charlie this week? What other forms of exercise did you do this week and for how many minutes?

[Your transcribed response will appear here]

Procedure

After obtaining informed consent, the participants completed a short survey about use of speech recognition technology and provided demographic information. They were then briefed that the tasks consisted of problem solving and recall and given instructions about how to use the speech recognition interface before commencing the experiment. Training continued until the participant clearly understood all aspects required to perform the experimental tasks; no practice tasks were undertaken. Participants completed two simple and two complex tasks, half of which were randomly assigned to using speech recognition (Figure 3). At the end of each task, cognitive load was assessed. To avoid order effects, the assignment of simple and complex tasks for the two modalities and task sequence were randomized. All voice recordings for the speech recognition tasks were captured independently to determine errors. At the end of all four tasks, the System Usability Scale (SUS) questionnaire [25] was completed for each modality followed by a feedback interview.

Figure 3. Experimental flow diagram.



Outcome Measures and Analysis

Participants' responses to the lifestyle management tasks with speech recognition and keyboard and mouse were compared using data extracted from the computer log, audio recordings of participant responses, and paper surveys.

Cognitive load was measured using a cognitive load inventory that was adapted from a validated instrument to reflect the nature of the tasks in the present study [26] (Multimedia Appendix 2). Self-ratings of using this inventory have shown it to be reliable, unobtrusive, and sensitive to small differences [23]. This inventory has been widely used [23,26,27], including in controlled studies of clinical decision making [28]. The inventory was administered on paper at the end of each condition.

Task completion time was measured in seconds and calculated from computer logs as the difference between the task start and end time.

The error rate for a given task was defined as the number of failed recall and problem-solving responses, calculated as a percentage of the maximum score for the task. Participants were asked to answer each of the conversational agent's questions correctly by recalling information or solving for numerical answers. Each task had predefined correct answers along with numbers associated with correct problem-solving and recall responses (Multimedia Appendix 1). For example, if the correct answers to a task were "line dancing" and "50 minutes," then the maximum score was 2. If a participant incorrectly recalled "line dancing" but correctly answered "50 minutes," then the error rate would be calculated as 50%. To ensure that transcription errors were not mistakenly attributed to participants, the voice recording on each response was used to validate the answers for tasks that were completed with speech recognition.

Usability was assessed using the SUS, a validated 10-item questionnaire [25], resulting in overall usability scores and scores of subelements (usability and learnability).

Statistical Analyses

The effects of speech recognition on cognitive load, task completion time, error rate, and SUS scores were tested using repeated-measures analysis to control for individual differences. The Wilcoxon signed-rank test was used because the results were not normally distributed. We compared speech recognition and keyboard and mouse across both levels of complexity. Cohen criteria were used to calculate and interpret effect sizes (*r*), where 0.1 indicates a small effect, 0.3 indicates a medium effect, and 0.5 indicates a large effect [29]. We estimated that a sample size of 42 was required to detect a difference of 25% and a standard deviation of 20% for each outcome measure for all tasks with 90% power and *P*<.05 [30]. No baseline measures could be derived from the literature. Descriptive statistics were used to summarize demographic information. All statistical analyses were undertaken using SPSS v24.0.0.0 software (IBM Corp, Armonk, NY, USA).

Results

Participants

Fifty-two university staff and students participated in the experiment. One participant did not complete the experiment and data from another was lost due to a technical error, leaving a total of 50 participants for inclusion in the analysis. The participants were aged 18-64 years, 30% (15/50) were 25-34 years and 54% (27/50) were women. More than half of the participants (56%, 28/50) reported never using speech recognition in their daily lives, 28% (14/50) reported using it once or twice a week, 14% (7/50) reported using it more than multiple times a week, and 4% (2/50) used speech recognition multiple times a day.

Effects of Speech Recognition on Cognitive Load

Participants experienced a significantly higher cognitive load when using speech recognition to perform the prescribed tasks (Table 1). These findings were consistent across both levels of task complexity, although the effect size increased from medium for simple tasks to high for complex tasks.

Table 1. Comparison of cognitive load between keyboard and mouse and speech recognition use by task complexity.

Task complexity	Keyboard and mouse, median (IQR)	Speech recognition, median (IQR)	*Z* value[a]	*P* value	Effect size (*r*)
Complex	3.2 (2.1-4.2)	5.2 (3.6-6.3)	–4.22	<.001	0.60
Simple	2.2 (2.2-4.4)	3.3 (2.2-5.6)	–2.24	.02	0.32

[a] Wicoxon signed-rank test.

Task Completion Time and Error Rate

Participants took significantly longer to complete tasks using speech recognition than the keyboard and mouse for complex tasks; however, there was no difference observed for simple tasks (Table 2). For complex tasks, there was a statistically significant increase in task completion time with a medium effect size (*r*=0.36); however, there was no difference for simple tasks.

There was no difference in error rates for both simple and complex tasks (Table 2). For complex tasks, we examined error rates by their type and found no difference for both problem solving (*Z*=–1.96, *P*=.05) and recall error rates (*Z*=–1.55, *P*=.12).

Table 2. Comparison of task completion time and error rates between keyboard and mouse and speech recognition use by task complexity.

Task complexity	Keyboard and mouse	Speech recognition	*Z* value[a]	*P* value
Complex, median (IQR)				
Completion time (seconds)	162 (124-192)	173 (136-223)	2.52	.01
Error rate (%)	11 (0-25)	11 (0-36)	1.80	.07
Simple, median (IQR)				
Completion time (seconds)	90 (74-124)	83 (68-11)	7.90	.43
Error rate (%)	0 (0-33)	0 (0-33)	–0.33	.74

[a] Wilcoxon signed-rank test.

Usability and Participant Perceptions About Speech Recognition

Participants found speech recognition to be significantly less usable than the keyboard and mouse. This was consistent with the SUS factor analysis, which revealed that speech recognition was perceived to be significantly less usable and harder to learn with medium and large effect sizes, respectively (Table 3).

Table 3. Usability (SUSa scores) of speech recognition compared to keyboard and mouse.

SUS category	Keyboard and mouse, median (IQR)	Speech recognition, median (IQR)	*Z* value[b]	*P* value	Effect size (*r*)
Overall	85 (72-90)	75 (62-85)	–3.30	.001	0.47
Usability	100 (97-100)	100 (75-100)	–2.98	.003	0.42
Learnability	81 (68-88)	72 (59-84)	–3.54	<.001	0.50

[a] SUS: System Usability Scale.

[b] Wilcoxon signed-rank test.

Overall, participants commented that the simulated conversational agent was “very straightforward” to use. Some reported that not having to think about spelling and grammar was an advantage of speech recognition. Although the participants found speech recognition to be a possibly more convenient way to enter free-text information, many observed that the major pitfalls of speech recognition were any transcription errors generated by the software, an inability to retract and edit answers once a sentence was uttered, and unwanted filler utterances such as “um” and “er” appearing in their responses. Participants also commented about the extra time taken to check and correct the output of speech recognition for transcribing errors when such errors could be prevented by using a keyboard and mouse in the first place. Others raised privacy issues as they were self-conscious about strangers eavesdropping on their responses.

When commenting on difficulties, some participants stated that they had struggled to formulate answers for scenarios and construct sentences to dictate to the conversational agent via speech recognition at the same time. They also commented that this action consumed more time to complete the task. By contrast, typing was perceived to be easier because it was “what we do every day.” This also allowed for using the textbox input to record answers as opposed to “keeping more in the brain” when storing information in cognitive, working memory. Speech was also perceived as a novelty because many were familiar with using a keyboard and mouse and saw it as a standard mode for human-computer interaction.

Discussion

Main Findings

When using speech recognition, participants reported a higher cognitive load for both simple and complex tasks compared to using a keyboard and mouse. Some participants reported an inability to think while speaking, which is consistent with previous observations [12,13]. A possible reason is due to an extra step introduced in processing the information and then having to form complete sentences before speaking. This extra processing step, which occurs once participants begin to formulate responses to the simulated conversational agent’s questions, may have involved working memory and contributed to participants experiencing a higher cognitive load while using speech recognition.

The higher cognitive load also provides a potential explanation for the more time required to complete complex tasks when using speech recognition and for it to be perceived less usable. Participants also noted a time delay required for speech recognition to process responses. It is thus possible that the extra time was introduced by the system and the time participants spent in checking the live transcription provided by speech recognition. If this was true, a consistent increase in time across all conditions would be expected. However, there was no significant time difference found for simple tasks. This could mean that the difference more likely arose from the task characteristics themselves. Further experiments are required to examine the interaction effect of modality and task complexity on cognitive load.

Despite previous reports about higher error rates with speech recognition for clinical documentation tasks [31], we found no difference in error rates between the two modalities for both simple and complex tasks. One possible reason that the second hypothesis was not supported is that most participants were observed to formulate their responses with caution by double-checking their answers when using speech recognition. For simple tasks, this meant looking at the scenario section that was on the same screen. For complex tasks, participants repeatedly clicked the “review scenario” button until they were satisfied with their answer.

Speech recognition was perceived to be less usable than a keyboard and mouse because it was harder to learn. A major factor contributing to this effect is that many participants reported being more comfortable with typing because it was an everyday human-computer interaction, especially in their profession. In addition, 56% of the participants did not use speech recognition in their daily lives, which possibly meant that they are less proficient in using speech recognition. Another possible factor is that the conversational agent was purposely designed to limit participants by not allowing users to correct transcription errors. This may be a common source of frustration that affects perceptions of usability for conversational agents. However, the error rate was unaffected because we used independent voice recordings to determine errors and the impact on usability was captured in the overall score.

Some participants reported lack of privacy as another major factor for speech recognition being less usable than a keyboard and mouse. A major advantage with typing was that a third party could not eavesdrop on interactions with the conversational agent. Thus, less confidence was placed on using the conversational agent with speech recognition compared with a keyboard and mouse. Although there was no difference in error rates, the increased cognitive load and task completion time may also have affected user experience.

Implications

Our findings suggest that speech recognition may not be uniformly suitable for the different contexts of health care. An important implication is for system designers to consider task characteristics and the resulting impact on cognitive load when selecting modalities for humans to interact with computers. In general, lower cognitive load, fewer errors, less time, and better usability are desirable. Speech recognition may be more suitable for frequent tasks such as documentation of notes in electronic health records (EHRs), which may generally not involve problem solving and recall (eg, when a doctor has already established a diagnosis during the patient consultation and uses the EHR to record their notes). However, it may not be suitable for occasional tasks such as incident reporting, which involves problem solving and recall, requiring clinicians to recollect the sequence of events and identify problems that led to an incident [32]. For such complex documentation tasks, the higher cognitive load and greater time required to use speech recognition, along with the lower usability, suggest that a keyboard and mouse may be a better input modality. By contrast, speech recognition may be necessary for contexts that require use of the hands and eyes.

One practical strategy for designers to assess the suitability of speech recognition as an input modality is to test user interfaces in the prototype stage with the cognitive load inventory, which is readily applicable to different modalities and systems. Importantly, designers need to consider privacy requirements when using speech recognition for busy health care settings, especially when sensitive health information is being handled. Our findings also suggest that privacy requirements may present a barrier for the use of speech recognition in some contexts such as health apps that require users to document their personal health information. These considerations will need to be evaluated on a case-by-case basis.

Comparison With the Literature

To the best of our knowledge, no previous studies have measured the effects of speech recognition in consumer digital health tasks. In clinical applications, the use of speech recognition for clinical documentation was found to increase error rates and task completion time in a controlled laboratory setting [31,33,34]. One possible reason for the disparity with our results may be the difference in the source of complexity. In the previous studies, complexity was distributed between the clinical scenario and the user interactions with an EHR, which required users to navigate to different sections of the record. By contrast, in our study, the type of user interaction remained the same for the different task types. The complexity instead arose from the scenario itself, which required users to problem solve and recall information from memory.

Limitations

There are several limitations to the design of the current study. We focused on the use of speech recognition as an input modality within a computer laboratory, which may not be representative of a real-world setting where environmental factors such as background noise and interruptions affect consumer interactions with digital health technologies. A general-purpose speech recognition engine that was not specifically optimized for the consumer health domain was tested on a desktop computer. This may have affected participant perceptions about usability and time required to use speech recognition. Participants were university students and staff, almost half of whom reported not using speech recognition technology in their daily lives. Therefore, our sample may not be representative of the general population of health consumers who might use conversational agents. However, because the participants were from a cohort that regularly used a keyboard and mouse, we were able to undertake a realistic assessment of the effort to learn and use speech recognition in the context of the conversational agent. For regular speech recognition users, expectations about the robustness and accuracy could have affected SUS scores. Although it is possible that individual differences such as health literacy, native language, pronunciation, fluency, and experience with speech recognition and a keyboard and mouse may have impacted the outcome variables, we attempted to control for these differences by using a within-subjects design. Further studies are needed to explore the influence of health and nutritional literacy. It is also possible that the quality of the speech recognition could have affected task completion time, but the effect would be consistent across experimental conditions. The error rate was unaffected because we used independent voice recordings to determine correct answers; therefore, the risk of the speech recognition mistranscribing speech by recording responses was controlled.

Despite these limitations, this study has contributed evidence related to the use of speech recognition as an input modality in human-computer interaction, particularly in a consumer digital health context. These results provide baseline measures of cognitive load in using speech recognition. Further studies using a more representative population of conversational agent users are needed to investigate the effects of cognitive load on task performance when speech recognition is integrated with consumer digital health technologies in real-world settings, including mobile devices such as smartphones and tablet computers.

Conclusions

This study found that using a keyboard and mouse was preferable to speech recognition for complex tasks involving problem solving and recall. This may be due to the higher cognitive load reported when using speech recognition and that the participants were more comfortable using a keyboard and mouse. Our results suggest that task characteristics need to be considered by designers when selecting the most appropriate input modality for human-computer interaction. Further studies using a broader variety of consumer digital health tasks of varying complexity are needed to investigate the contexts in which use of speech recognition is appropriate. The effects of cognitive load on task performance and its significance also need to be investigated.

Acknowledgments

We thank the participants who gave up their time to participate in this study. We thank Vitaliy Kim and Owen Sun for sharing their expertise to build the simulated conversational agent.

Authors' Contributions

FM conceptualized the study design. JC conducted the research and primary analysis, and drafted the paper. LL helped to formulate the scenarios and DL helped to adapt the cognitive load inventory. JC is responsible for the integrity of the work as the guarantor of the study. All authors participated in writing and revising the paper. All aspects of the study (including design; collection, analysis, and interpretation of data; writing of the report; and decision to publish) were led by the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tasks.
[DOCX File , 22 KB - jmir_v22i6e14827_app1.docx]

Multimedia Appendix 2
Cognitive load inventory.
[DOCX File , 16 KB - jmir_v22i6e14827_app2.docx]

References

1. Hinton G, Deng L, Yu D, Dahl G, Mohamed A, Jaitly N, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Process Mag 2012 Nov;29(6):82-97. [doi: 10.1109/MSP.2012.2205597]
2. Hodgson T, Magrabi F, Coiera E. Evaluating the usability of speech recognition to create clinical documentation using a commercial electronic health record. Int J Med Inform 2018 May;113:38-42. [doi: 10.1016/j.ijmedinf.2018.02.011] [Medline: 29602431]
3. Kumah-Crystal Y, Pirtle C, Whyte H, Goode E, Anders S, Lehmann C. Electronic Health Record Interactions through Voice: A Review. Appl Clin Inform 2018 Jul 18;9(3):541-552 [FREE Full text] [doi: 10.1055/s-0038-1666844] [Medline: 30040113]
4. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
5. Black L, McTear M, Black N, Harper R, Lemon M. Appraisal of a conversational artefact and its utility in remote patient monitoring. 2005 Presented at: 18th IEEE Symposium on Computer-Based Medical Systems; June 23-24, 2005; Dublin. [doi: 10.1109/cbms.2005.33]
6. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. J Med Internet Res 2006 Dec 11;8(4):e30 [FREE Full text] [doi: 10.2196/jmir.8.4.e30] [Medline: 17213048]
7. Beveridge M, Fox J. Automatic generation of spoken dialogue from medical plans and ontologies. J Biomed Inform 2006 Oct;39(5):482-499 [FREE Full text] [doi: 10.1016/j.jbi.2005.12.008] [Medline: 16495159]
8. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
9. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. J Am Med Inform Assoc 2017 Mar 01;24(2):246-250. [doi: 10.1093/jamia/ocw154] [Medline: 28011595]
10. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
11. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
12. Shneiderman B. The limits of speech recognition. Commun ACM 2000;43(9):63-65. [doi: 10.1145/348941.348990]

13. Ruan S, Wobbrock JO, Liou K, Ng A, Landay JA. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Jan 08;1(4):1-23 [FREE Full text] [doi: 10.1145/3161187]
14. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci 2001 Feb;24(1):87-114. [doi: 10.1017/s0140525x01003922] [Medline: 11515286]
15. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956;63(2):81-97. [doi: 10.1037/h0043158]
16. Peterson L, Peterson MJ. Short-term retention of individual verbal items. J Exp Psychol 1959;58(3):193-198. [doi: 10.1037/h0049234] [Medline: 14432252]
17. Sweller J. Cognitive load during problem solving: effects on learning. Cognitive Sci 1988;12(2):257-285. [doi: 10.1207/s15516709cog1202_4]
18. Limerick H, Moore J, Coyle D, editors. Empirical evidence for a diminished sense of agency in speech interfaces. 2015 Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; : ACM; 2015; Seoul p. 3967-3970. [doi: 10.1145/2702123.2702379]
19. Leitner G, Ahlström D, Hitz M. Usability of Mobile Computing in Emergency Response Systems – Lessons Learned and Future Directions. 2007 Nov Presented at: Symposium of the Austrian HCI and Usability Engineering Group; November 22, 2007; Graz. [doi: 10.1007/978-3-540-76805-0_20]
20. Sweller J. Cognitive load theory, learning difficulty, and instructional design. Learn Instruct 1994 Jan;4(4):295-312. [doi: 10.1016/0959-4752(94)90003-5]
21. Ayres P, Sweller J. Locus of Difficulty in Multistage Mathematics Problems. Am J Psychol 1990;103(2):167-193. [doi: 10.2307/1423141]
22. Ayres P. Systematic Mathematical Errors and Cognitive Load. Contemp Educ Psychol 2001 Apr;26(2):227-248. [doi: 10.1006/ceps.2000.1051] [Medline: 11273658]
23. Sweller J, Ayres P, Kalyuga S. Intrinsic and Extraneous Cognitive Load. In: Cognitive Load Theory. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies Vol. 1. New York: Springer; 2011:57-69.
24. Beverloo P, Bringert B, Capiel G, Carter J, Kharidi N, Mazzoni D. W3C Community. 2012. Web Speech API Specification URL: https://www.w3.org/community/speech-api/ [accessed 2018-03-03]
25. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. London: Taylor & Francis; 1996:189-194.
26. Paas FG. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. J Education Psychol 1992;84(4):429-434. [doi: 10.1037/0022-0663.84.4.429]
27. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Education Psychol 2003 Mar;38(1):63-71. [doi: 10.1207/S15326985EP3801_8]
28. Lyell D, Magrabi F, Coiera E. The Effect of Cognitive Load and Task Complexity on Automation Bias in Electronic Prescribing. Hum Factors 2018 Nov 25;60(7):1008-1021. [doi: 10.1177/0018720818781224] [Medline: 29939764]
29. Pallant J. Non-parametric statistics. In: SPSS Survival Manual, 3rd ed. Adelaide: McGraw-Hill Education; 2010:210-231.
30. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
31. Hodgson T, Magrabi F, Coiera E. Evaluating the Efficiency and Safety of Speech Recognition within a Commercial Electronic Health Record System: A Replication Study. Appl Clin Inform 2018 Apr 16;9(2):326-335 [FREE Full text] [doi: 10.1055/s-0038-1649509] [Medline: 29768633]
32. Sun O, Chen J, Magrabi F. Using Voice-Activated Conversational Interfaces for Reporting Patient Safety Incidents: A Technical Feasibility and Pilot Usability Study. Stud Health Technol Inform 2018;252:139-144. [Medline: 30040696]
33. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. J Am Med Inform Assoc 2016 Apr 17;23(e1):e169-e179 [FREE Full text] [doi: 10.1093/jamia/ocv152] [Medline: 26578226]
34. Hodgson T, Magrabi F, Coiera E. Efficiency and safety of speech recognition for documentation in the electronic health record. J Am Med Inform Assoc 2017 Nov 01;24(6):1127-1133. [doi: 10.1093/jamia/ocx073] [Medline: 29016971]

Abbreviations

EHR: electronic health records
SUS: System Usability Scale

Edited by G Eysenbach; submitted 28.05.19; peer-reviewed by M Lee, G Zuccon, L Kelly; comments to author 03.10.19; revised version received 27.11.19; accepted 29.03.20; published 01.06.20.

Please cite as:
Chen J, Lyell D, Laranjo L, Magrabi F
Effect of Speech Recognition on Problem Solving and Recall in Consumer Digital Health Tasks: Controlled Laboratory Experiment
J Med Internet Res 2020;22(6):e14827
URL: https://www.jmir.org/2020/6/e14827
doi:10.2196/14827
PMID:32442129

Review

Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review

Alaa Abd-Alrazaq[1], BSc, MSc, PhD; Zeineb Safi[1], BSc, MSc; Mohannad Alajlani[2], BSc, MSc, MBA, PhD; Jim Warren[3], BS, PhD; Mowafa Househ[1], BCom, MEng, PhD; Kerstin Denecke[4], Dr rer nat

[1]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[2]Institute of Digital Healthcare, University of Warwick, Coventry, United Kingdom

[3]School of Computer Science, University of Auckland, Auckland, New Zealand

[4]Institute for Medical Informatics, Bern University of Applied Sciences, Bern, Switzerland

Corresponding Author:
Kerstin Denecke, Dr rer nat
Institute for Medical Informatics
Bern University of Applied Sciences
Quellgasse 21
2502 Biel
Bern
Switzerland
Phone: 41 76 409 97 61
Email: kerstin.denecke@bfh.ch

Abstract

Background: Dialog agents (chatbots) have a long history of application in health care, where they have been used for tasks such as supporting patient self-management and providing counseling. Their use is expected to grow with increasing demands on health systems and improving artificial intelligence (AI) capability. Approaches to the evaluation of health care chatbots, however, appear to be diverse and haphazard, resulting in a potential barrier to the advancement of the field.

Objective: This study aims to identify the technical (nonclinical) metrics used by previous studies to evaluate health care chatbots.

Methods: Studies were identified by searching 7 bibliographic databases (eg, MEDLINE and PsycINFO) in addition to conducting backward and forward reference list checking of the included studies and relevant reviews. The studies were independently selected by two reviewers who then extracted data from the included studies. Extracted data were synthesized narratively by grouping the identified metrics into categories based on the aspect of chatbots that the metrics evaluated.

Results: Of the 1498 citations retrieved, 65 studies were included in this review. Chatbots were evaluated using 27 technical metrics, which were related to chatbots as a whole (eg, usability, classifier performance, speed), response generation (eg, comprehensibility, realism, repetitiveness), response understanding (eg, chatbot understanding as assessed by users, word error rate, concept error rate), and esthetics (eg, appearance of the virtual agent, background color, and content).

Conclusions: The technical metrics of health chatbot studies were diverse, with survey designs and global usability metrics dominating. The lack of standardization and paucity of objective measures make it difficult to compare the performance of health chatbots and could inhibit advancement of the field. We suggest that researchers more frequently include metrics computed from conversation logs. In addition, we recommend the development of a framework of technical metrics with recommendations for specific circumstances for their inclusion in chatbot studies.

(J Med Internet Res 2020;22(6):e18301) doi:10.2196/18301

KEYWORDS

chatbots; conversational agents; health care; evaluation; metrics

Introduction

Background

The potential of human-computer dialog to provide health care benefits has been perceived for many decades. In 1966, Weizenbaum's ELIZA system caught the public imagination with its imitation of a psychotherapist through the relatively simple linguistic token manipulation possible at the time [1]. From the mid-1990s, a family of interventions based on automated telephone sessions (telephone-linked care) demonstrated effectiveness in promoting health adherence across a range of behaviors including medication, diet, and physical activity [2]. As mobile phones have become commonplace, a range of SMS text messaging–based interventions have been developed and trialed, with particular success in smoking cessation [3]. At the same time, internet/web-based interventions have shown the ability to promote positive health behavior change [4,5], and the interaction components associated with users sticking with an internet intervention are increasingly well understood and include the inclusion of dialog elements [6].

With the advent of smartphones, the distribution of highly interactive chatbots has been greatly facilitated, particularly with the ubiquitous use of app stores and wide installation of chat apps that can include chatbots, notably Facebook Messenger. Chatbots, as with other electronic health (eHealth) interventions, offer scalability and 24-hour availability to plug gaps in unmet health needs. For example, Woebot delivers cognitive behavior therapy and has been tested with students with depression [7]. The students who used Woebot significantly reduced their symptoms of depression over the study period as measured by the depression questionnaire PHQ-9, while those in the information control group (who instead read a self-help book) did not [7]. In recent years, artificial intelligence (AI) based on deep learning has created waves with its ability to outperform physicians at some diagnostic tasks [8,9]. XiaoIce is a social chatbot that emphasizes emotional connection and it has communicated with over 660 million active users since its launch in 2014 [10]; its performance shows that deep learning can be successfully applied to meaningful dialog tasks. Combining the factors of ease-of-distribution, successful applications, and AI methods to improve health chatbot performance, it is reasonable to expect health chatbots in increasing numbers and variety to take an increasingly serious role in the health care system.

Research Problem and Aim

To be an evidence-based discipline requires measurement of performance. The impact of health chatbots on clinical outcomes is the ultimate measure of success. For example, did the condition (eg, depression, diabetes) improve to a statistically significant degree on an accepted measure (eg, PHQ-9 [11] or hemoglobin A1c [12], respectively), as compared to a control group? Such studies, however, may require large sample sizes to detect the effect and provide relatively little insight into the mechanism by which the chatbot achieves the change; additionally, studies may provide particularly little insight if the result is negative.

As an alternative and useful precursor to clinical outcome metrics, technical metrics concern the performance of the chatbot itself (eg, did participants feel that it was usable, give appropriate responses, and understand their input?). Appropriateness refers to the relevance of the provided information in addressing the problem prompted [13]. Furthermore, this includes more objective measures of the chatbot interaction, such as the number of conversational turns taken in a session or time taken, and measures that require some interpretation but are still well-defined, such as task completion. These technical measures offer a potential method for comparison of health chatbots and for understanding the use and performance of a chatbot to decide if it is working well enough to warrant the time and expense of a trial to measure clinical outcomes.

Previously, we had introduced a framework for evaluation measures of health chatbots to provide guidance to developers [14]. The framework development, however, was based on a relatively informal process vulnerable to the authors' biases in terms of what studies were considered in its formulation. Therefore, the aim of this study is to use a rigorous review methodology to identify the technical metrics used by previous studies to evaluate health care chatbots. The final goal of these efforts is to be able to make recommendations for an evaluation framework for health chatbots.

Methods

Overview

To achieve the aforementioned objective, a scoping review was conducted. To conduct a transparent and replicable review, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Extension for Scoping Reviews (PRISMA-ScR) guidelines [15]. This research was conducted by an interdisciplinary team of researchers with backgrounds in nursing, computer science, and mental health applications.

Search Strategy

Search Sources

For this review, we searched the following bibliographic databases November 1-3, 2019: MEDLINE (via EBSCO), EMBASE (Excerpta Medica Database; via Ovid), PsycINFO (via Ovid), CINAHL (Cumulative Index of Nursing and Allied Health Literature; via EBSCO), IEEE (Institute of Electrical and Electronics Engineers) Xplore, ACM (Association for Computing Machinery) Digital Library, and Google Scholar. We screened only the first 100 hits retrieved by searching Google Scholar, as it usually retrieves several thousand references ordered by their relevance to the search topic. We checked the reference list of the included studies to identify further studies relevant to the current review (ie, backward reference list checking). Additionally, we used the "cited by" function available in Google Scholar to find and screen studies that cited the included studies (ie, forward reference list checking).

Search Terms

The search terms were derived from previously published literature and the opinions of informatics experts. For health-related databases, we used search terms related to the intervention of interest (eg, chatbot, conversational agent, and chat-bot). In addition to intervention-related terms, we used terms related to the context (eg, health, disease, and medical) for non–health-related databases (eg, IEEE and ACM digital library). Multimedia Appendix 1 details the search strings used for searching each electronic database.

Study Eligibility Criteria

The intervention of interest in this review was chatbots that are aimed at delivering health care services to patients. Chatbots implemented in stand-alone software or web-based platforms were included. However, we excluded chatbots operated by a human (Wizard-of-Oz) or integrated into robotics, serious games, SMS text messaging, or telephone systems. To be included, studies had to report a technical evaluation of a chatbot (eg, usability, classifier performance, and word error rate). We included peer-reviewed articles, dissertations, and conference proceedings, and we excluded reviews, proposals, editorials, and conference abstracts. This review included studies written in the English language only. No restrictions were considered regarding the study design, study setting, year of publication, and country of publication.

Study Selection

Authors MA and ZS independently screened the titles and abstracts of all retrieved references and then independently read the full texts of included studies. Any disagreements between the two reviewers were resolved by AA. We assessed the intercoder agreement by calculating Cohen, which was 0.82 for screening titles and abstracts and 0.91 for reading full texts, indicating a very good agreement [16].

Data Extraction

To conduct a reliable and accurate extraction of data from the included studies, a data extraction form was developed and piloted using 8 included studies (Multimedia Appendix 2). The data extraction process was independently conducted by two reviewers (MA and ZS) and a third reviewer (AA) resolved any disagreements. Intercoder agreement between the reviewers was good (Cohen κ=0.67).

Data Synthesis

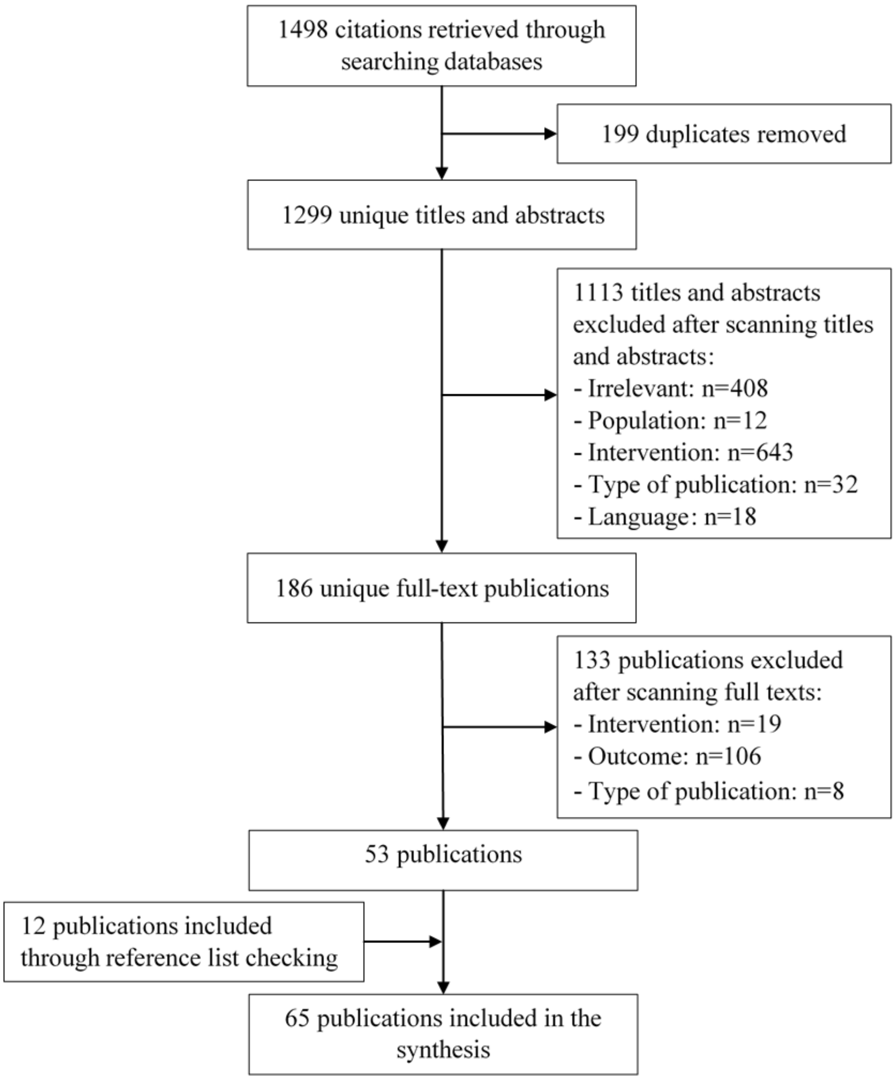
A narrative approach was used to synthesize the extracted data. After identifying all technical metrics used by the included studies to evaluate chatbots, we divided them into 4 categories based on the aspect of chatbots that the metrics evaluate. The 4 categories were formed after a discussion by the authors in which consensus was reached. For each metric, we identified how the studies measured it. Data synthesis was managed using Microsoft Excel (Microsoft Corporation).

Results

Search Results

By searching the 7 electronic databases, 1498 citations were retrieved. After removing 199 (13.3%) duplicates of these citations, 1299 (86.7%) titles and abstracts were screened. The screening process resulted in excluding 1113 (74.3%) titles and abstracts due to several reasons detailed in Figure 1. When we read the full text of the remaining 186 (12.4%) citations, a further 133 (8.9%) citations were excluded (Figure 1). In total, 12 studies were found by backward and forward reference checking. We included 65 studies in the review.

Figure 1. Flowchart of the study selection process.



Description of Included Studies

Characteristics of the included studies are detailed in Table 1. Cross-sectional survey was the most commonly used study design (n=41, 63%). About 57% (n=37) of the included studies were published as journal articles. Half of the studies (n=33, 51%) were conducted in the United States. Approximately 70% (n=45) of the studies were published between 2015 and 2019.

The sample size was reported in 61 studies, and 38 studies (62%) had 50 or fewer participants. In 44 studies, the age of participants was reported; the mean age of participants was 39 years, with a range of 13-79 years. Sex of participants was reported in 54 studies, where the mean percentage of males was 48.1%. Of the 62 studies that reported participants' health conditions, 34 (54.8%) studies recruited participants from a clinical population (ie, those with health issues). Participants were recruited from clinical settings (n=30, 49.2%), community (n=20, 32.8%), and educational settings (n=18, 29.5%). Metadata and population characteristics of each included study are presented in Multimedia Appendix 3.

Chatbots were used for self-management (n=17, 26.2%), therapeutic purposes (n=12, 18.5%), counselling (n=12, 18.5%), education (n=10, 15.4%), screening (n=9, 13.8%), training (n=7, 10.8%), and diagnosing (n=3, 4.6%). Although chatbots were implemented in stand-alone software in about 62% (n=40) of studies, chatbots were implemented in web-based platforms in the remaining studies (n=25, 39%). Chatbot responses were generated based on predefined rules, machine learning approaches, or both methods (hybrid) in 82% (N=53), 17% (n=11), and 2% (n=1) of the included studies, respectively. In the majority of studies (n=58, 89%), chatbots led the dialogue. In about 62% (n=40) of studies, users interacted with chatbots only by typing in their utterances (texts). The most common modalities used by chatbots to interact with users were a combination of text, voice, and nonverbal language (ie, facial expression and body language; n=21, 32%), text only (n=20, 31%), and a combination of voice and nonverbal language (n=19, 29%). The most common disorders targeted by chatbots were any health condition (n=20, 31%) and depression (n=15, 23%). Multimedia Appendix 4 displays characteristics of the intervention in each included study.

Table 1. Characteristics of the included studies (N=65).

Parameters and characteristics	Studies, n (%)[a]
Study metadata	
Study design	
Survey	41 (63)
Quasi-experiment	11 (17)
Randomized controlled trial	13 (20)
Type of publication	
Journal article	37 (57)
Conference proceeding	25 (39)
Thesis	3 (5)
Country	
United States	33 (51)
France	5 (8)
Netherlands	3 (5)
Japan	3 (5)
Australia	3 (5)
Italy	2 (3)
Switzerland and Netherlands	2 (3)
Finland	1 (2)
Sweden	1 (2)
Turkey	1 (2)
United Kingdom	1 (2)
Switzerland & Germany	1 (2)
Mexico	1 (2)
Spain	1 (2)
Global population	1 (2)
Romania, Spain and Scotland	1 (2)
Philippines	1 (2)
Switzerland	1 (2)
New Zealand	1 (2)
Spain and New Zealand	1 (2)
South Africa	1 (2)
Year of publication	
Before 2010	3 (5)
2010-2014	17 (26)
2015-2019	45 (70)
Population characteristics	
Sample size[b]	
≤50	38 (62)
51-100	11 (18)
101-200	9 (15)
>200	3 (5)

Parameters and characteristics	Studies, n (%)[a]
Age (years)[c]	
Mean (range)	39 (13-79)
Sex (%)[e]	
Male	48.1
Health condition[f]	
Clinical sample	34 (55)
Nonclinical sample	28 (45)
Setting[g,h]	
Clinical	30 (50)
Community	20 (33)
Educational	18 (30)
Intervention characteristics	
Purpose[i]	
Self-management	17 (26)
Therapy	12 (19)
Counselling	12 (19)
Education	10 (15)
Screening	9 (14)
Training	7 (11)
Diagnosing	3 (5)
Platform	
Stand-alone software	40 (62)
Web-based	25 (39)
Response generation	
Rule-based	53 (82)
Artificial intelligence	11 (17)
Hybrid	1 (2)
Dialogue initiative	
Chatbot	58 (89)
Users	4 (6)
Both	3 (5)
Input modality	
Text	40 (62)
Voice	9 (14)
Voice and nonverbal	8 (12)
Text and voice	6 (9)
Text and nonverbal	2 (3)
Output modality	
Text, voice and nonverbal	21 (32)
Text	20 (31)
Voice and nonverbal	19 (29)

Parameters and characteristics	Studies, n (%)[a]
Text & voice	4 (6)
Voice	1 (2)
Targeted disorders[j]	
Any health condition	20 (31)
Depression	15 (23)
Autism	5 (8)
Anxiety	5 (8)
Substance use disorder	5 (8)
Posttraumatic stress disorder	5 (8)
Mental disorders	3 (5)
Sexually transmitted diseases	3 (5)
Sleep disorders	2 (3)
Diabetes	2 (3)
Alzheimer	1 (2)
Asthma	1 (2)
Cervical cancer	1 (2)
Dementia	1 (2)
Schizophrenia	1 (2)
Stress	1 (2)
Genetic variants	1 (2)
Cognitive impairment	1 (2)
Atrial fibrillation	1 (2)

[a] Percentages were rounded and may not sum to 100.

[b] Sample size was reported in 61 studies.

[c] Mean age was reported in 44 studies.

[d] N/A: not applicable.

[e] Sex was reported in 54 studies.

[f] Sample type was reported in 62 studies.

[g] Setting was reported in 61 studies.

[h] Numbers do not add up as several chatbots focused on more than one health condition.

[i] Numbers do not add up as several chatbots have more than one purpose.

[j] Numbers do not add up as several chatbots target more than one health condition.

Results of Studies

Overview

The included studies evaluated chatbots using many technical metrics, which were categorized into 4 main groups: metrics related to chatbots as a whole (global metrics), metrics related to response generation, metrics related to response understanding, and metrics related to esthetics. More details about metrics are presented in the following sections.

Global Metrics

The included studies evaluated chatbots as a whole using the following metrics: usability, classifier performance, speed, technical issues, intelligence, task completion rate, dialogue efficiency, dialogue handling, context awareness, and error management.

Usability of chatbots was assessed in 37 (56.9%) studies [17-53]. Usability was evaluated using a single question in a self-administrated questionnaire [17,20-25,29-31,33,34,36,37,40,42,44,45,47-51,53], multiple questions in a self-administrated questionnaire [28,41,43], a specific questionnaire (eg, system usability scale [SUS] questionnaire) [18,26,27,32,35,38,39,46,52], or observation [19].

Classifier performance of chatbots was evaluated in 8 (12.3%) studies [54-61]. Many metrics were used to measure the classifier performance, namely: area under curve [54,55,60,61], accuracy [56-58,61], sensitivity [55,57,59,60], specificity

[55,57,59,60], positive predictive value [55,57,60], and negative predictive value [55,60]. The speed of chatbots was examined in 4 studies [29,35,53,62]. The speed was evaluated using a single question in a self-administrated questionnaire [29,35], multiple questions in a self-administrated questionnaire [53], and interviews [62].

Technical issues (eg, errors/glitches) in chatbots were examined in 4 studies (6.2%) [7,36,51,63]. Technical issues were assessed through interviews [7,51,63], a single question in a self-administrated questionnaire [36], and checking staff logs [51]. In addition, 3 studies assessed the intelligence of chatbots using either multiple questions in a self-administrated questionnaire [41,64] or a single question in a self-administrated questionnaire [27]. In 3 studies, the task completion rate was examined by checking the conversation logs [38,53,65].

Of the reviewed studies, 2 (3.1%) studies examined chatbot flexibility in dialogue handling (eg, its ability to maintain a conversation and deal with users' generic questions or responses that require more, less, or different information than was requested) using interviews [27] and multiple questions in a self-administrated questionnaire [38]. Dialogue efficiency of chatbots, which refers to the number of dialogue steps required to finish a task, was assessed in 1 study by reviewing transcribed conversation logs [38]. The same study examined the chatbot's context awareness (ie, its ability to utilize contextual knowledge to appropriately respond to users) using multiple questions in a self-administrated questionnaire [38]. Error management, which refers to a chatbot's ability to detect and understand misspelled words in users' replies (eg, "anious" instead of anxious), was examined in only 1 study [27].

Metrics Related to Response Generation

The following metrics were utilized by the included studies to evaluate response generation by chatbots: appropriateness of responses, comprehensibility, realism, speed of response, empathy, repetitiveness, clarity of speech, and linguistic accuracy.

Of the reviewed studies, 15 (23.1%) examined the appropriateness and adequacy of verbal [18,19,27,28,31,34,38,39,51,58,66-69] and nonverbal responses of chatbots [32]. Appropriateness of responses was assessed using interviews [18,19,31,34,51,66,68], a single question in self-administrated questionnaire [27,32,39,67], conversation logs [38,58,69], and multiple questions in self-administrated questionnaire [28].

Comprehensibility of responses, which refers to the degree to which a chatbot generates responses understandable by users, was evaluated by 14 (21.5%) studies [20,23,31,34,36,39,42,44,51,52,59,60,63,69]. Comprehensibility of responses was evaluated using a single question in a self-administrated questionnaire [20,23,31,36,39,42,44,52,59,60,63,69] and interviews [34,51].

In total, 14 (21.5%) studies assessed how human-like chatbots are (realism) [17,18,21,34,39,41,46,50,63,66,68,70-72]. Realism of chatbots was examined in terms of verbal responses only [17,21,34,39,46,63,68,70], nonverbal responses only [66], or both verbal and nonverbal responses [18,41,50,71,72]. The included studies evaluated realism using a single question in a self-administrated questionnaire [17,18,21,39,46,50,63,70], multiple questions in a self-administrated questionnaire [41,72], and interviews [18,34,66,68,71].

Altogether, 11 (16.9%) studies assessed the speed of a chatbot's responses [18,19,28,30,34,36,38,68-70,73]. The speed of responses was examined using a single question in a self-administrated questionnaire [18,30,36,69,70,73], interviews [19,34,68], multiple questions in a self-administrated questionnaire [53], and system logs [38]. Empathy of a chatbot, which refers to its ability to show empathy to users, was examined in 10 studies [7,35,41,42,64,66,67,71,73,74]. Those studies evaluated empathy using a single question in a self-administrated questionnaire [7,35,41,42,67,71,73], interviews [66,71], and multiple questions in a self-administrated questionnaire [64].

Repetitiveness of a chatbot's responses was examined in 9 (13.8%) studies [7,20,27,53,57,66,73,75,76]. Repetitiveness of responses was evaluated using a single question in a self-administrated questionnaire [7,20,27,53,57,73] and interviews [66,75,76]. We found that 6 (9.2%) studies evaluated clarity or quality of speech using either interviews [51,62,75] or a single question in a self-administrated questionnaire [27,69,77]. Linguistic accuracy of a chatbot's responses was evaluated in 2 (3.7%) studies using a single question in a self-administrated questionnaire [31,35].

Metrics Related to Response Understanding

The included studies evaluated chatbot understanding of users' responses using the following metrics: understanding as assessed by users, word error rate, concept error rate, and attention estimator errors.

Chatbot understanding, which refers to a chatbot's ability to adequately understand the verbal and nonverbal responses of users, was assessed by 20 (30.8%) studies [7,18,20,23,27,32,33,36,39,41,42,53,57,59,63,64,68,73,78,79]. Of those studies, 2 studies assessed understanding of both verbal and nonverbal responses [18,79], 1 study assessed understanding of nonverbal responses only [32], and the remaining studies assessed understanding of verbal responses only. The understanding of responses was evaluated using multiple questions in a self-administrated questionnaire in 4 studies [42,64,78,79], interviews in 2 studies [18,68], and a single question in a self-administrated questionnaire in the remaining studies.

Word error rate, which assesses the performance of speech recognition in chatbots, was examined in 2 (3.7%) studies using conversational logs [65,69]. Concept error rate, which depends on the correct recognition of the semantic result of a user utterance, was evaluated in 1 study by checking conversational logs [65]. Attention estimation, which refers to a chatbot's ability to determine whether the user is gazing toward the screen or away from it, was examined in 1 study by checking conversational logs [69].

Metrics Related to Esthetics

The included studies evaluated the esthetics of chatbots using the following metrics: appearance of the virtual agent, background color and content, font type and size, button color, shape, icon, and background color contrast.

In total, 5 (7.7%) studies assessed the appearance of the virtual agent using a single question in a self-administrated questionnaire [69,77,80], interviews [51], and focus groups [45]. In addition, 1 (1.5%) study evaluated background color, color contrast, and content; font type and size; and button color, shape, and icon using a survey [80].

Discussion

Principal Findings

It became clear that there is currently no standard method in use to evaluate health chatbots. Most aspects are studied using self-administered questionnaires or user interviews. Common metrics are response speed, word error rate, concept error rate, dialogue efficiency, attention estimation, and task completion. Various studies assessed different aspects of chatbots, complicating direct comparison. Although some of this variation may be due to the individual characteristics of chatbot implementations and their distinct use cases, it is difficult to see why metrics such as appropriateness of responses, comprehensibility, realism, speed of response, empathy and repetitiveness are each only applicable to a small percentage of cases. Further, objective quantitative metrics (eg, those based on log reviews) were comparatively rarely used in the reported studies. We thus suggest continuing research and development toward an evaluation framework for technical metrics with recommendations for specific circumstances for their inclusion in chatbot studies.

Jadeja et al [81] introduced 4 dimensions for chatbot evaluations: the information retrieval (IR) perspective, the user experience (UX) perspective, the linguistic perspective, and the AI (human-likeness) perspective. In earlier work [14], we adapted and broadened this categorization, modifying the IR perspective to a task-oriented perspective since health chatbots are not necessarily designed only to retrieve information; additionally, we included system quality and health care quality perspectives. Excluding the health care quality perspective, which is outside the definition of technical metrics, the findings of this scoping review show that all these dimensions are indeed represented in health chatbot evaluations. Rather, the issue appears to be the inconsistency in what is measured and how, along with the skew toward self-reporting and the UX perspective. Additional work is still required to come up with standard metrics and corresponding assessment tools specifically addressing quality in health chatbots.

We found usability to be the most commonly assessed aspect of health chatbots. The system usability scale (SUS [82,83]) is a well-established usability instrument that we observed was used repeatedly, although it was not used in the majority of the studies assessing usability; in many cases, a single survey question was used instead. The SUS is nonproprietary, technology-agnostic, and designed to support comparison across products [82]. As such, global assessment of the user experience of health chatbots could be enhanced in quality and comparability by researchers standardizing on inclusion of the SUS in their evaluations. However, studies by Holmes et al [84] showed that conventional methods for assessing usability and user experience may not be as accurate when applied to health chatbots. As such, there remains research to be done toward appropriate metrics for health chatbots.

Conversational-turns per session (CPS) has been suggested as a success metric for social chatbots as exemplified by XiaoIce [85]. Although the aims for health chatbots are not identical to those of social chatbots, if CPS gains acceptance as a standard measure in the social chatbot domain, it would make it a leading candidate for a standard measure to include in health chatbot evaluations to assess their social engagement dimension. An alternative or supplementary measure related to the social dimension would be to have users score the chatbot on empathy; however, CPS has the advantage of being an objective and quantitative measure. Other objective and quantitative measures such as interaction time or time on task could be alternatives to CPS, but might be less representative of engagement than CPS if for instance the user was multitasking chatbot interaction with other tasks. Beyond social engagement, task completion (often assessed by analyzing conversation logs) is another promising global measure.

A further area for standardization would be in the quality of responses. We observed response generation to be widely measured but in very diverse ways. Emergence of standard measures for response generation and understanding would greatly advance the comparability of studies. Development of validated instruments in this area would be a useful contribution to chatbot research.

We commend the inclusion of classifier performance in health chatbot studies where this is applicable and practical to assess. It could be less meaningful to compare raw performance (eg, as area under the curve) across domains due to differences in difficulty; ideally, chatbot performance would be compared to the performance of a human expert for the task at hand. Further, we perceive the opportunity for a progression of performance measures in health chatbot studies as a given product gains maturity. Good early-stage metrics would be those that assess response quality and response understanding to establish that the product is working well. Subsequent experiments can advance the assessment of self-reported usability and metrics of social engagement. Where applicable, classifier performance can round out technical performance evaluation to establish whether trials to assess clinical outcomes are warranted.

Strengths and Limitations

Strengths

This study is the first review that summaries the technical metrics used by previous studies to evaluate health care chatbots. This helps readers explore how chatbots were evaluated in health care. Given that this review was executed and reported in line with PRISMA-ScR guidelines [1], it could be considered a high-quality review.

To retrieve as many relevant studies as possible, the most commonly used databases in the fields of health and information technology were searched. Further, we searched Google Scholar and conducted backward and forward reference list checking to retrieve gray literature and minimize the risk of publication bias.

As two reviewers independently selected the studies and extracted the data, the selection bias in this review was minimal. This review can be considered a comprehensive review given that we did not apply restrictions regarding the study design, study setting, year of publication, and country of publication.

Laranjo et al [86] reviewed the characteristics, current applications and evaluation measures of health chatbots. In contrast to our work, they did not solely concentrate on the technical metrics used for chatbot evaluations. The metrics they reviewed included task completion or word accuracy. In contrast to Laranjo et al [86], who included only 17 papers reporting on 14 different conversational agents, our work is more comprehensive as it included 65 publications. Further, we had a different research question in mind while conducting the review.

Limitations

This review focused on chatbots that are aimed at delivering health care services to patients and work on stand-alone software and web browsers; it excluded chatbots that used robotics, serious games, SMS text messaging, Wizard-of-Oz, and telephones. Thus, this review did not include many technical metrics used to evaluate chatbots for other users (eg, physicians, nurses, and caregivers), in other fields (eg, business and education), or with alternative modes of delivery (eg, SMS text messaging, Wizard-of-Oz, and telephones). The abovementioned restrictions were applied by previous reviews about chatbots as these features are not part of ordinary chatbots [87-90].

Due to practical constraints, we could not search interdisciplinary databases (eg, Web of Science and ProQuest), conduct a manual search, or contact experts. Further, the search in this review was restricted to English-language studies. Accordingly, it is likely that this review missed some studies.

Conclusion

From this review, we perceive the need for health chatbot evaluators to consider measurements across a range of aspects in any given study or study series, including usability, social experience, response quality, and, where applicable, classifier performance. The establishment of standard measures would greatly enhance comparability across studies with the SUS and CPS as leading candidates for usability and social experience, respectively. It would be ideal to develop guidelines for health chatbot evaluators indicating what should be measured and at what stages in product development. Development of validated measurement instruments in this domain is sparse and such instruments would benefit the field, especially for response quality metrics.

Authors' Contributions

AA developed the protocol and conducted the search with guidance from and under the supervision of KD and MH. Study selection and data extraction were carried out independently by MA and ZS. AA solved any disagreements between the two reviewers. AA synthesized the data. AA and KD drafted the manuscript, and it was revised critically for important intellectual content by all authors. KD and JW reviewed the related literature and interpreted the results. All authors approved the manuscript for publication and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string.
[DOCX File , 19 KB - jmir_v22i6e18301_app1.docx]

Multimedia Appendix 2
Data extraction form.
[DOCX File , 17 KB - jmir_v22i6e18301_app2.docx]

Multimedia Appendix 3
Metadata and population characteristics of each included study.
[DOCX File , 24 KB - jmir_v22i6e18301_app3.docx]

Multimedia Appendix 4
Characteristics of the intervention in each included study.
[DOCX File , 23 KB - jmir_v22i6e18301_app4.docx]

References

1. Weizenbaum J. ELIZA-a computer program for the study of natural language communication between man and machine. Commun ACM 1966;9(1):36-45. [doi: 10.1145/365153.365168]
2. Friedman R. Automated Telephone Conversations to Assess Health Behavior and Deliver Behavioral Interventions. Journal of Medical Systems 1998 Apr 01;22(2):95-102. [doi: 10.1023/A:1022695119046]
3. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
4. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
5. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
6. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
7. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
8. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
9. Stead WW. Clinical Implications and Challenges of Artificial Intelligence and Deep Learning. JAMA 2018 Sep 18;320(11):1107-1108. [doi: 10.1001/jama.2018.11029] [Medline: 30178025]
10. Zhou L, Gao J, Li D, Shum H. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. Computational Linguistics 2020 Mar;46(1):53-93. [doi: 10.1162/coli_a_00368]
11. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
12. Conlin PR, Colburn J, Aron D, Pries RM, Tschanz MP, Pogach L. Synopsis of the 2017 U.S. Department of Veterans Affairs/U.S. Department of Defense Clinical Practice Guideline: Management of Type 2 Diabetes Mellitus. Ann Intern Med 2017 Nov 07;167(9):655-663. [doi: 10.7326/M17-1362] [Medline: 29059687]
13. Kocaballi A, Quiroz J, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E. Responses of Conversational Agents to Health and Lifestyle Prompts: Investigation of Appropriateness and Presentation Structures. J Med Internet Res 2020;22(2):e15823. [doi: 10.2196/preprints.15823]
14. Denecke K, Warren J. How to Evaluate Health Applications with Conversational User Interface? Studies in Health Technology and Informatics 2020:1-9.
15. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Altman D. Practical statistics for medical research. London: Chapman & Hall/CRC; 1990.
17. Abdullah AS, Gaehde S, Bickmore T. A Tablet Based Embodied Conversational Agent to Promote Smoking Cessation among Veterans: A Feasibility Study. J Epidemiol Glob Health 2018 Dec;8(3-4):225-230. [doi: 10.2991/j.jegh.2018.08.104] [Medline: 30864768]
18. Ali M, Rasazi Z, Mamun A, Langevin R, Rawassizadeh R, Schubert L. A Virtual Conversational Agent for Teens with Autism: Experimental Results and Design Lessons. arXiv 2018:1-13.
19. Beiley M. Mental Health And Wellness Chatbot. United States: The University of Arizona; 2019.
20. Bickmore T, Caruso L, Clough-Gorr K. Acceptance and usability of a relational agent interface by urban older adults. In: CHI'05 extended abstracts on Human factors in computing systems: ACM. 2005 Presented at: CHI EA '05; April; Portland, OR, USA p. 1212-1215. [doi: 10.1145/1056808.1056879]
21. Bickmore TW, Pfeifer L, Jack BW. Taking the time to care. 2009 Presented at: Proceedings of the 27th international conference on Human factors in computing systems - CHI 09; 2009; Boston. [doi: 10.1145/1518701.1518891]
22. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
23. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15 Suppl 2:197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
24. Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: Relational agents for antipsychotic medication adherence. Interacting with Computers 2010 Jul;22(4):276-288. [doi: 10.1016/j.intcom.2010.02.001]

25. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
26. Bresó A, Martínez-Miranda J, Botella C, Baños RM, García-Gómez JM. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. Expert Systems 2016 May 25;33(4):297-312. [doi: 10.1111/EXSY.12151] [Medline: 2016]
27. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S. Assessing the Usability of a Chatbot for Mental Health Care. Internet Science 2019;Cham:121-132. [doi: 10.1007/978-3-030-17705-8_11]
28. Comendador BEV, Francisco BMB, Medenilla JS, Nacion SMT, Serac TBE. Pharmabot: A Pediatric Generic Medicine Consultant Chatbot. JOACE 2015;3(2):137-140. [doi: 10.12720/joace.3.2.137-140]
29. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
30. Demirci H. User experience over time with conversational agents case study of woebot on supporting subjective well-being. Turkey: Middle East Technical University; 2018.
31. Denecke K, Hochreutener S, Pöpel A, May R. Self-Anamnesis with a Conversational User Interface: Concept and Usability Study. Methods Inf Med 2019 Mar 15;57(05/06):243-252. [doi: 10.1055/s-0038-1675822]
32. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A, et al. SimSensei kiosk: a virtual human interviewer for healthcare decision support. 2014 Presented at: International Foundation for Autonomous Agents and Multiagent Systems; May 2014; Paris p. 1061-1068.
33. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, Feasibility, and Preliminary Efficacy of a Theory-Based Relational Embodied Conversational Agent Mobile Phone Intervention to Promote HIV Medication Adherence in Young HIV-Positive African American MSM. AIDS Educ Prev 2019 Feb;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
34. Elmasri D, Maeder A. A Conversational Agent for an Online Mental Health Intervention. In: Brain Informatics and Health: International Conference. 2016 Presented at: Brain Informatics and Health 2016; October 13-16; Omaha, NE, USA p. 243-251. [doi: 10.1007/978-3-319-47103-7_24]
35. Fadhil A, Schiavo G, Villafiorita A, Fondazione B. OlloBot-Towards A Text-Based Arabic Health Conversational Agentvaluation and Results. 2019 Presented at: Proceedings of Recent Advances in Natural Language Processing; 2019; Varna p. 295-303. [doi: 10.26615/978-954-452-056-4_034]
36. Griol D, Callejas Z. Mobile Conversational Agents for Context-Aware Care Applications. Cogn Comput 2015 Aug 21;8(2):336-356. [doi: 10.1007/s12559-015-9352-x]
37. Hanke S, Sandner E, Kadyrov S, Stainer-Hochgatterer A. Daily life support at home through a virtual support partner. 2016 Presented at: 2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL); 2016; London p. 1-7. [doi: 10.1049/ic.2016.0058]
38. Hess G, Fricker G, Denecke K. Improving and Evaluating eMMA?s Communication Skills: A Chatbot for Managing Medication. Studies in health technology and informatics 2019;259:101-104.
39. Kadariya D, Venkataramanan R, Yip H, Kalra M, Thirunarayanan K, Sheth A. kBot: Knowledge-Enabled Personalized Chatbot for Asthma Self-Management. 2019 Presented at: IEEE International Conference on Smart Computing (SMARTCOMP); 2019; Washington p. 138-143. [doi: 10.1109/smartcomp.2019.00043]
40. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. IT - Information Technology 2019;61(5-6):253-263. [doi: 10.1515/itit-2019-0019]
41. Lisetti C, Amini R, Yasavur U, Rishe N. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. ACM Trans Manage Inf Syst 2013 Dec 01;4(4):1-28. [doi: 10.1145/2544103]
42. Magnani JW, Schlusser CL, Kimani E, Rollman BL, Paasche-Orlow MK, Bickmore TW. The Atrial Fibrillation Health Literacy Information Technology System: Pilot Assessment. JMIR Cardio 2017;1(2):e7 [FREE Full text] [doi: 10.2196/cardio.8543] [Medline: 29473644]
43. Micoulaud-Franchi J, Sagaspe P, de SE, Bioulac S, Sauteraud A, Philip P. Acceptability of Embodied Conversational Agent in a Health Care Context. Intelligent Virtual Agents 2016:416-419. [doi: 10.1007/978-3-319-47665-0_45]
44. Milne M, Luerssen M, Lewis T, Leibbrandt R, Powers D. Development of a virtual agent based social tutor for children with autism spectrum disorders. The 2010 International Joint Conference on Neural Networks (IJCNN). 2010 Presented at: The 2010 International Joint Conference on Neural Networks (IJCNN); 2010 July 2010; Barcelona p. 18-23. [doi: 10.1109/ijcnn.2010.5596584]
45. Schmidlen T, Schwartz M, DiLoreto K, Kirchner HL, Sturm AC. Patient assessment of chatbots for the scalable delivery of genetic counseling. J Genet Couns 2019 Dec;28(6):1166-1177. [doi: 10.1002/jgc4.1169] [Medline: 31549758]
46. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. Canada: ACM; 2018 Presented at: The 2018 CHI Conference on Human Factors in Computing Systems; 2018; Montreal p. 1-15. [doi: 10.1145/3173574.3173972]

47. Smith MJ, Ginger EJ, Wright M, Wright K, Boteler Humm L, Olsen D, et al. Virtual reality job interview training for individuals with psychiatric disabilities. J Nerv Ment Dis 2014 Sep;202(9):659-667 [FREE Full text] [doi: 10.1097/NMD.0000000000000187] [Medline: 25099298]
48. Smith MJ, Ginger EJ, Wright K, Wright MA, Taylor JL, Humm LB, et al. Virtual reality job interview training in adults with autism spectrum disorder. J Autism Dev Disord 2014 Oct;44(10):2450-2463 [FREE Full text] [doi: 10.1007/s10803-014-2113-y] [Medline: 24803366]
49. Smith MJ, Humm LB, Fleming MF, Jordan N, Wright MA, Ginger EJ, et al. Virtual Reality Job Interview Training for Veterans with Posttraumatic Stress Disorder. J Vocat Rehabil 2015;42(3):271-279 [FREE Full text] [doi: 10.3233/JVR-150748] [Medline: 27721645]
50. Tanaka H, Sakti S, Neubig G, Toda T, Negoro H, Iwasaka H. Automated Social Skills Trainer. 2015 Presented at: The 20th International Conference on Intelligent User Interfaces; 2015; Atlanta. [doi: 10.1145/2678025.2701368]
51. Thompson D, Callender C, Gonynor C, Cullen KW, Redondo MJ, Butler A, et al. Using Relational Agents to Promote Family Communication Around Type 1 Diabetes Self-Management in the Diabetes Family Teamwork Online Intervention: Longitudinal Pilot Study. J Med Internet Res 2019 Sep 13;21(9):e15318 [FREE Full text] [doi: 10.2196/15318] [Medline: 31538940]
52. Tielman ML, Neerincx MA, Bidarra R, Kybartas B, Brinkman W. A Therapy System for Post-Traumatic Stress Disorder Using a Virtual Agent and Virtual Storytelling to Reconstruct Traumatic Memories. J Med Syst 2017 Aug;41(8):125 [FREE Full text] [doi: 10.1007/s10916-017-0771-y] [Medline: 28699083]
53. Yasavur U, Lisetti C, Rishe N. Let's talk! speaking virtual counselor offers you a brief intervention. J Multimodal User Interfaces 2014 Sep 5;8(4):381-398. [doi: 10.1007/s12193-014-0169-9]
54. Amato F, Marrone S, Moscato V, Piantadosi G, Picariello A, Sansone C. Chatbots Meet eHealth: Automatizing Healthcare. 2017 Presented at: Workshop on Artificial Intelligence with Application in Health; 2017; Bari p. 1-10.
55. Auriacombe M, Moriceau S, Serre F, Denis C, Micoulaud-Franchi J, de Sevin E, et al. Development and validation of a virtual agent to screen tobacco and alcohol use disorders. Drug and Alcohol Dependence 2018 Dec;193(6):1-6. [doi: 10.1016/J.drugalcdep.2018.08.025] [Medline: 2018]
56. Ghosh S, Bhatia S, Bhatia A. Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System. Stud Health Technol Inform 2018;252:51-56. [Medline: 30040682]
57. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR mHealth uHealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
58. Ni L, Lu C, Liu N, Liu J. MANDY: Towards a Smart Primary Care Chatbot Application. Knowledge and Systems Sciences 2017:38-52. [doi: 10.1007/978-981-10-6989-5_4]
59. Philip P, Bioulac S, Sauteraud A, Chaufton C, Olive J. Could a Virtual Human Be Used to Explore Excessive Daytime Sleepiness in Patients? Presence: Teleoperators and Virtual Environments 2014 Nov;23(4):369-376. [doi: 10.1162/pres_a_00197]
60. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7:42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
61. Tanaka H, Adachi H, Ukita N, Ikeda M, Kazui H, Kudo T, et al. Detecting Dementia Through Interactive Computer Avatars. IEEE J Transl Eng Health Med 2017;5:1-11. [doi: 10.1109/jtehm.2017.2752152]
62. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729. [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
63. Razavi S, Ali M, Smith T, Schubert L, Hoque M. The LISSA virtual human and ASD teens: An overview of initial experiments. 2016 Presented at: International Conference on Intelligent Virtual Agents; 2016; Los Angeles p. 460-463. [doi: 10.1007/978-3-319-47665-0_55]
64. Liu B, Sundar SS. Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
65. Turunen M, Hakulinen J, Ståhl O, Gambäck B, Hansen P, Rodríguez Gancedo MC, et al. Multimodal and mobile conversational Health and Fitness Companions. Computer Speech & Language 2011 Apr;25(2):192-209. [doi: 10.1016/j.csl.2010.04.004]
66. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
67. Swartout W, Artstein R, Forbell E, Foutz S, Lane HC, Lange B, et al. Virtual Humans for Learning. AIMag 2013 Dec 15;34(4):13. [doi: 10.1609/aimag.v34i4.2487]
68. van Heerden A, Ntinga X, Vilakazi K. The potential of conversational agents to provide a rapid HIV counseling and testing services. 2017 Presented at: International Conference on the Frontiers and Advances in Data Science; 2017; Xi'an p. 80-85. [doi: 10.1109/fads.2017.8253198]

69. Wargnier P, Benveniste S, Jouvelot P, Rigaud A. Usability assessment of interaction management support in LOUISE, an ECA-based user interface for elders with cognitive impairment. TAD 2018 Nov 26;30(3):105-126. [doi: 10.3233/tad-180189]
70. Kang J, Wei L. “Give Me the Support I Want!”: The Effect of Matching an Embodied Conversational Agent’s Social Support to Users’ Social Support Needs in Fostering Positive User-Agent Interaction. 2018 Presented at: The 6th International Conference on Human-Agent Interaction; 2018; Southampton p. 106-113. [doi: 10.1145/3284432.3284462]
71. Olafsson S, O'Leary T, Bickmore T. Coerced Change-talk with Conversational Agents Promotes Confidence in Behavior Change. 2019 Presented at: The 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; 2019; Trento p. 31-40. [doi: 10.1145/3329189.3329202]
72. Tielman ML, Neerincx MA, van Meggelen M, Franken I, Brinkman W. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. Technol Health Care 2017 Dec 04;25(6):1081-1096 [FREE Full text] [doi: 10.3233/THC-170899] [Medline: 28800346]
73. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
74. Martínez-Miranda J, Bresó A, García-Gómez J. Look on the bright side: a model of cognitive change in virtual agents. 2014 Presented at: International Conference on Intelligent Virtual Agents; 2014; Boston p. 285-294. [doi: 10.1007/978-3-319-09767-1_37]
75. Bickmore T. Relational agents affecting change through human-computer relationships. USA: Massachusetts Institute of Technology; 2003:1-284.
76. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interventions 2017 Dec;10:39-46. [doi: 10.1016/j.invent.2017.10.002]
77. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. J Med Syst 2019 Jun 25;43(8):246. [doi: 10.1007/s10916-019-1387-1] [Medline: 31240494]
78. Pinto MD, Greenblatt AM, Hickman RL, Rice HM, Thomas TL, Clochesy JM. Assessing the Critical Parameters of eSMART-MH: A Promising Avatar-Based Digital Therapeutic Intervention to Reduce Depressive Symptoms. Perspect Psychiatr Care 2016 Jul;52(3):157-168. [doi: 10.1111/ppc.12112] [Medline: 25800698]
79. Yokotani K, Takagi G, Wakashima K. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. Computers in Human Behavior 2018 Aug;85(6):135-145. [doi: 10.1016/j.chb.2018.03.045] [Medline: 2018]
80. Wu Y, Samant D, Squibbs K, Chaet A, Morshedi B, Barnes L. Design of Interactive Cancer Education Technology for Latina Farmworkers. 2014 Apr Presented at: The 2014 IEEE International Symposium on Software Reliability Engineering Workshops; 2014; Washington URL: http://europepmc.org/abstract/MED/29978858 [doi: 10.1109/SIEDS.2014.6829908]
81. Jadeja M, Varia N. Perspectives for evaluating conversational AI. arXiv preprint 2017:1-6.
82. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
83. Brooke J. SUS-A quick and dirty usability scale. In: Jordan P, Thomas B, McClelland I, Weerdmeester B, editors. Usability evaluation in industry. London: CRC Press; 1996:4-7.
84. Holmes S, Moorhead A, Bond R, Zheng H, Coates V, McTear M. Can we use conventional methods to assess conversational user interfaces? Usability testing of a healthcare chatbot; 2019 Presented at: The 31st European Conference on Cognitive Ergonomics; 2019; Belfast p. 207-2014. [doi: 10.1145/3335082.3335094]
85. Shum H, He X, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers Inf Technol Electronic Eng 2018 Jan 8;19(1):10-26. [doi: 10.1631/fitee.1700826]
86. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
87. Abd-alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: A scoping review. International Journal of Medical Informatics 2019 Dec;132:103978. [doi: 10.1016/j.ijmedinf.2019.103978]
88. Abd-alrazaq A, Rababeh A, Alajlani M, Bewick B, Househ M. The effectiveness and safety of using chatbots to improve mental health: A systematic review and meta-analysis. Journal of Medical Internet Research. Forthcoming 2020:2020 (forthcoming). [doi: 10.2196/16021]
89. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
90. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]

Abbreviations

ACM: Association for Computing Machinery
AI: artificial intelligence
CINAHL: Cumulative Index of Nursing and Allied Health Literature
CPS: conversational-turns per session
eHealth: electronic health
EMBASE: Excerpta Medica Database
IEEE: Institute of Electrical and Electronics Engineers
IR: information retrieval
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses-Extension for Scoping Reviews
UX: user experience

Edited by C Lovis; submitted 18.02.20; peer-reviewed by DU Iqbal, E Bellei; comments to author 08.04.20; revised version received 13.04.20; accepted 15.04.20; published 05.06.20.

Please cite as:
Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K
Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review
J Med Internet Res 2020;22(6):e18301
URL: http://www.jmir.org/2020/6/e18301/
doi:10.2196/18301
PMID:32442157

Original Paper

Using Blockchain Technology to Mitigate Challenges in Service Access for the Homeless and Data Exchange Between Providers: Qualitative Study

Anjum Khurshid[1*], MBBS, PhD; Vivian Rajeswaren[2*], BSc; Steven Andrews[1*], PhD

[1]Dell Medical School, The University of Texas at Austin, Austin, TX, United States
[2]University of Colorado at Denver, Anshutz Medical Campus, Aurora, CO, United States
[*]all authors contributed equally

Corresponding Author:
Anjum Khurshid, MBBS, PhD
Dell Medical School
The University of Texas at Austin
1701 Trinity Street
Austin, TX, 78712
United States
Phone: 1 5124955225
Email: anjum.khurshid@austin.utexas.edu

Abstract

Background: In the homeless population, barriers to housing and supportive services include a lack of control or access to data. Disparate data formats and storage across multiple organizations hinder up-to-date intersystem access to records and a unified view of an individual's health and documentation history. The utility of blockchain to solve interoperability in health care is supported in recent literature, but the technology has yet to be tested in real-life conditions encompassing the complex regulatory standards in the health sector.

Objective: This study aimed to test the feasibility and performance of a blockchain system in a homeless community to securely store and share data across a system of providers in the health care ecosystem.

Methods: We performed a series of platform demonstrations and open-ended qualitative feedback interviews to determine the key needs and barriers to user and stakeholder adoption. Account creation and data transactions promoting organizational efficiency and improved health outcomes in this population were tested with homeless users and service providers.

Results: Persons experiencing homelessness and care organizations could successfully create accounts, grant and revoke data sharing permissions, and transmit documents across a distributed network of providers. However, there were issues regarding the security of shared data, user experience and adoption, and organizational preparedness for service providers as end users. We tested a set of assumptions related to these problems within the project time frame and contractual obligations with an existing blockchain-based platform.

Conclusions: Blockchain technology provides decentralized data sharing, validation, immutability, traceability, and integration. These core features enable a secure system for the management and distribution of sensitive information. This study presents a concrete evaluation of the effectiveness of blockchain through an existing platform while revealing limitations from the perspectives of user adoption, cost-effectiveness, scalability, and regulatory frameworks.

(J Med Internet Res 2020;22(6):e16887) doi:10.2196/16887

KEYWORDS

blockchain; distributed ledger technology; health care; data sharing; homeless; data autonomy

Introduction

Background

According to Austin's Ending Community Homelessness Coalition (ECHO), more than 7100 people in Austin experienced homelessness and needed housing and other supportive services in 2016 [1]. One of the many barriers to housing for persons experiencing homelessness (PEH) is the lack of access to their own documentation and the time and money required to replace lost or stolen documents. Client data are often distributed between service providers, causing a deterioration in health and delays in service when accessing data from multiple sources. Existing database systems such as Homeless Management Information System (HMIS) to manage documentation for the PEH lack interoperability for different document formats and organizations outside the system, do not facilitate data ownership for clients, and cannot guarantee data privacy and security.

The Mayor's Challenge Competition, sponsored by Bloomberg Philanthropies, is designed to facilitate innovative and scalable solutions for city leaders to tackle urgent local challenges. Austin, Texas, was one of the 35 cities selected to prototype and test their ideas over 6 months with a grant support of US $100,000. Through a partnership between Dell Medical School, the City of Austin government, Austin Travis County Emergency Medical Services, and community organizations, a pilot study was conducted to test the first use of blockchain technology to validate identities and improve access to services for the PEH in the country.

Objectives

To solve these problems of *transaction identity*, we previously proposed the use of distributed ledger technology (DLT), or blockchain technology, to provide a validated, immutable identity and record of service transactions [2]. The literature shows promise using blockchain for data control in electronic health records, public health surveillance, disease management, genomic and biomedical data, and clinical trials [3-9]. Blockchain is currently used in rural Indonesia, Syrian refugee camps, and the slums of Kenya for similar identity management applications. Key life events such as birth registration, immunizations, health provider visits, and agricultural harvests are securely recorded and maintained on a blockchain, creating an immutable, accessible, and portable identity for displaced and marginalized people. Following this model, we explored the use of a true *economic passport* for the homeless, working toward the goal of ending homelessness. This technology was chosen as a solution over other methods of storing and sharing information such as a digital wallet or a relational database because its features met the requirements of our use case. We have continued this study, and here, we provide results from our testing of the deployment of this approach.

Blockchain is a technology for shared databases used by multiple writers in which each organization's entries must be verified and cannot be modified by an outside party. When transactions from one entity are needed by another to provide a service or verify a document, storing transactions in a single shared database promotes expedient exchange of information. A regular shared database can reliably manage the permissioning of multiple authorized users to view data. However, a relational database system in which various entities update and write information can result in unrecoverable errors. Maintenance of a health and social service information system involves backup storage services, recovery mechanisms, and updating information [7]. In the event of a database server failure, the entire system is affected, and information can be lost if it has not been backed up and stored. Ownership of the master file by a centralized user also accords control to a single party; thus, there is no inherent safeguard against data tampering, and the integrity of data cannot be guaranteed [10].

Blockchains organize data so that secure transactions are approved and recorded through consensus from entities on the chain [10,11], providing greater error checking and transactional validity than relational shared databases [10-12]. A blockchain distributes data across the network, with data copied on each node of the chain [13]. Each node installs the genesis block or the first block in the chain [10]. A group of validated transactions is added to a new block with the file attachment, sender, receiver, timestamp, and cryptographic hash of the previous block [14]. A hash is a one-way encryption function, which is used to generate a public and private key for each user [15]. Information sent over the blockchain is secured by a user's private key and cannot be viewed or modified without the key [12,13]. Data are encrypted and unintelligible to protect private information stored on the blockchain or in the event of a security breach [14]. An individual block's hash depends on the hash of the previous block, locking transactions together [12,14]. Modifying data in one block would alter all subsequent blocks, making the blockchain an immutable and authoritative record of transactions [10,12-14]. Blockchain also uses a consensus validation mechanism replacing a trusted third-party intermediary or a manual offline reconciliation process with peer-to-peer protocols, allowing organizations to agree upon submitted entries without a singular point of failure or control [10-12,16]. A network of users collectively adheres to previously agreed upon rules automatically implemented to verify the authenticity of transactions and ordering of records added to the chain [12,16].

A *public* blockchain is completely decentralized, and transactions depend on consensus from a majority of nodes. In a *private* blockchain, users are granted access by permission from the owner of the blockchain [10]. For sensitive data such as personal identity records, private blockchains limit transaction visibility to authorized users and promote scalability because of greater user and transactional control [12]. A private blockchain is only partially decentralized because nodes are limited to trusted users with varying degrees of access and sharing permissions. If a conflict or security breach arises, the system can be recovered from any user and timestamp [10]. The only security issue is that a majority of nodes could collude to rewrite the chain, as there is a partially centralized authority controlling the nodes in the chain. However, in a permissioned system, it is unlikely that users aim to violate the immutability of the blockchain, as they are trusted entities using the blockchain for organizational or personal benefit [12,16-18].

Lack of interoperability between service providers arises from nonuniform data formats and storage methods [13]. Blockchain creates an accessible and authoritative ledger of diverse document types, acting as a method for storing and gathering information from multiple independent systems [6]. Off-blockchain data storage can be integrated for large files or extensive data storage in a variety of formats and is scalable to include a larger health and social service ecosystem. The blockchain can also be used to facilitate communication for application programming interfaces (APIs), which can restructure, aggregate, and merge data from various sources in a standardized format. A blockchain-based API model allows for decentralized and authoritative data exchange between systems, user identification, a validated transaction history, and proof of transaction legitimacy [19].

In this study, we describe the results of a pilot test using a blockchain solution to mitigate the current challenges in service access for the homeless and data exchange between providers. The potential for blockchain in various health care settings has been examined in several studies [3,4,6-9,14,20,21]. However, these studies are theoretical in nature, and to our knowledge, no published studies have examined the feasibility, effectiveness, or performance of blockchain in a real-life public setting. In this study, we used a private, permissioned blockchain system for secure storage and transmission of documents with planned API integration for intersystem data access and transmission from legacy databases. This model additionally permits individuals to access and control their own data by uploading official documents and sharing self- or provider-uploaded records with select organizations. Our approach in this study aimed to increase agency, motivation, and control while reducing service barriers for the homeless and supplying more complete and accurate information for service providers.

Methods

Generating Testable Assumptions

Austin, Texas, was 1 of the 35 cities selected by the 2018 Mayor’s Challenge Competition, sponsored by Bloomberg Philanthropies. The competition was designed to facilitate innovative and scalable solutions for city leaders to tackle urgent local challenges. Each Champion City selected in the Mayor’s Challenge Competition was to prototype and test their ideas over 6 months. Through a partnership between Dell Medical School, the City of Austin’s Office of Technology Innovation, Austin Travis County Emergency Medical Services, and community organizations, a pilot study was conducted to test the use of blockchain technology to validate identities for the homeless.

Participation in the Mayor’s Challenge Competition required a series of steps to design and implement testing. First, we identified assumptions underlying the idea, which will be required for stakeholder participation. We found that the following set of assumptions must be tested and shown to be valid for the successful implementation of our solution (see Textbox 1): (1) our prototype platform would be appropriate for PEH in Austin, (2) PEH will be able to understand its features and will consent to participate in such a system, (3) providers will be able to access and trust the information on the platform, (4) blockchain technology is essential to manage identity for PEH, (5) privacy and confidentiality of data will be protected, (6) service delivery will be facilitated, and (7) interorganizational efficiency will increase by sharing data in a standardized manner with shared governance on the platform. After determining how to test and evaluate these assumptions, we developed and modified prototypes and tests based on real-life findings, user feedback, and an evolving understanding of our idea. For prototype testing, we used an existing blockchain technology platform that had been implemented outside the United States to manage the identities of refugees.

Second, we identified markers for what constitutes an appropriate test of a platform based on blockchain technology (see Table 1). To be effective, the platform needs to be able to allow individuals to enter the system by creating an account. They need to be able to upload documents and then share those documents with others, with control over which documents get shared with which other participants. Service providers need to be able to additionally conduct transactions, especially transactions of identity validation, and to share information with each other. All these actions and pieces of information need to be recorded and available, to be differentially accessible based on user-controlled permissions, and to remain securely protected and immutable. The system overall needs to be easily understood by users.

Before testing our assumptions, initial engagement with the homeless population was performed to determine their needs and concerns. Two-hour meetings were held biweekly over a period of 12 months. Attendees included 20 homeless individuals in the City of Austin and 1 to 2 staff members from the City of Austin Office of Innovation who regularly interacted with the homeless population. The homeless participants were rotated monthly to include a larger subset of the population, and each participant was compensated for their time at a rate of US $20 per hour. Meetings consisted of open-ended questions regarding difficulties those with lived experience of homelessness face in daily life, interacting with service providers, maintaining documents, and filling out applications.

On the basis of these discussions and the answers we received to the questions in Table 2, we developed a list of documents that are most useful for testing our solution. Figure 1 lists types of documents and types of organizations and gives a sense of how difficult it is to reobtain a particular document. The figure provides insight into what a person experiencing homelessness needs to be able to manage to prove their identity to receive services, especially those that provide health care and potential housing.

As can be seen, the documents required for identity verification varied across types of organizations. Every organization, though, needed an official photo ID; yet, about one-third of clients lacked such a basic identity document when first entering the system. Replacing a photo ID is a time-intensive effort. Approximately half of the presenting users lacked insurance cards, which were required by many organizations. Looking at the entirety of the figure, though, it becomes clear that managing identity involves managing a significant number of identity documents across a wide audience of service organizations.

As some of the assumptions had overlapping aspects, we describe our methodology in testing all 7 assumptions in the subgroups below.

Textbox 1. Testable assumptions for solution implementation.

1. Given that our blockchain platform is modified from a pre-existing platform for refugees, the use case of refugees must be nominally equivalent to the use case of people experiencing homelessness
2. To gain user participation, people experiencing homelessness must understand the functionality and features of the platform, and consent to participate.
3. Stakeholder participation requires providers to access, accept, and trust information shared through the blockchain network.
4. Blockchain is necessary to meet our goals.
5. Privacy and confidentiality of personal data needs to be protected.
6. Enabling access to information facilitates service transactions.
7. The technology enables a shared data standard and governance that increases inter-organizational efficiency.

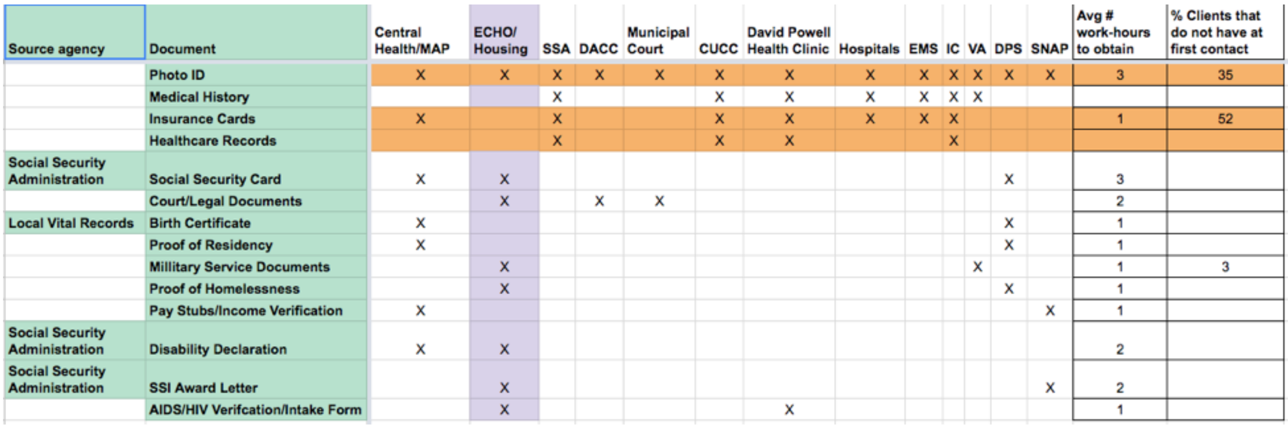
Table 1. Blockchain components needed for a successful prototype.

Test	Elements needed
(1) Allow individuals to create accounts and upload documents	• Profiles • Ability to connect profiles • Individual document repository • Ability to grant repository and asset level access
(2) Allow individuals to share documents with differentiated permissions	• Ability to find other profiles in the system • Ability to send a document • Ability to share a document • Profile user-controlled permission • Data level permission
(3) Allow service providers to share documents and conduct transactions, including verification of service use	• Ability to see a validated document • Ability to share a document • Ability to see permissions by individuals
(4) Allow service providers to share with each other	• Ability to find other verifiable users/profiles
(5) Capture all transactions/records	• Ledger with varied permissions • All actions write to a ledger • Inability to delete from ledger

Table 2. Research questions.

Particpants	Questions
For persons experiencing homelessness:	• Which things on your to-do list frustrate you the most? • Which records or documents are you in most need of accessing? • If you could reduce a barrier to a daily goal, what would it be?
For community health paramedics, when you help the homeless:	• Where do you see the most missed opportunities? • What common miscommunications could be prevented if you had direct information or document access? • What information is lacking that would help you the most?

Figure 1. Documents needed across service providers and their homeless clients. CUCC: Community Care Clinics; DACC: Downtown Austin Community Court ; DPS: Department of Public Safety; ECHO: Ending Community Homelessness Coalition; EMS: Emergency Medical Services; IC: Integral Care; SNAP: Supplemental Nutrition Assistance Program; SSA: Social Security Administration; VA: Veterans' Administration.



Source agency	Document	Central Health/MAP	ECHO/ Housing	SSA	DACC	Municipal Court	CUCC	David Powell Health Clinic	Hospitals	EMS	IC	VA	DPS	SNAP	Avg # work-hours to obtain	% Clients that do not have at first contact
	Photo ID	X	X	X	X	X	X	X	X	X	X	X	X	X	3	35
	Medical History			X			X	X	X	X	X	X				
	Insurance Cards	X		X			X	X	X	X	X				1	52
	Healthcare Records			X			X	X			X					
Social Security Administration	Social Security Card	X	X										X		3	
	Court/Legal Documents		X		X	X									2	
Local Vital Records	Birth Certificate	X											X		1	
	Proof of Residency	X											X		1	
	Millitary Service Documents		X									X			1	3
	Proof of Homelessness		X										X		1	
	Pay Stubs/Income Verification	X												X	1	
Social Security Administration	Disability Declaration	X	X												2	
Social Security Administration	SSI Award Letter		X											X	2	
	AIDS/HIV Verifcation/Intake Form		X					X							1	

Testing Assumptions 1 and 5: Usability and Security of the Platform

Assumptions 1 and 5 regarding the viability of an existing private, permissioned blockchain platform for our use case and the security and privacy of data were tested in 2 sessions on June 14, 2018, and July 13, 2018. In the first session, our testing plan was for the homeless clients to set up an account, upload a document onto the blockchain, deny a request to connect from an organization, accept a request to connect from an organization, and participate in a 3-party transaction between the client and 2 service providers. All tests were performed with platform provider staff, our team members, and 2 homeless individuals. Members of our team created mock provider accounts before testing with our homeless participants. For our second testing session, we modified the user interface so that an account profile contained the user's picture and name with categories for general profile information, identity documents, medical documents, and dependents or emergency contacts.

Testing Assumption 2: User Participation and Understanding of the Technology

We tested our assumption regarding user participation (assumption 2) at 2 pop-up resource clinics, where the homeless could access multiple health, social, and other service interventions. Our group of 8 city officials and 5 individuals from our team engaged a total of 34 homeless residents. Of the 34 homeless residents, 15 agreed to participate, 11 indicated interest when more information and a working platform were available, and 8 either declined to participate or did not finish the discussion.

Attendees were offered the opportunity to discuss and provide feedback on a technology designed to securely store and share their documents. In our discussions with homeless participants, we used 3 different prototypes of our platform with increasing levels of fidelity to a real-life DLT platform. At the lowest level, to explain the concept of DLT in a nontechnical manner, we created a prototype of the platform using several journals with a lock and key (Figure 2). Each journal represented the record of a user or a service provider. In a DLT, when a document (or other record) is written in the journal or uploaded onto the blockchain, none of the other users in the individual's network can see its contents until permission is granted, even if they are aware that there is an existing entry in the ledger (or block on the chain). In our testing analogy, granting permission is represented by the key that opens the lock to a specific physical journal. We also demonstrated distributed data storage by tearing up a piece of paper from a user's journal and spreading the pieces into buckets representing different institutions. In this scenario, when a document was accessed with permission, the pieces were reassembled.

Increasing the fidelity of the prototypes, we next showed screenshots of the platform prototype with sample transactions, and then, at the highest fidelity level, a digital prototype on a mobile phone with uploaded data was presented. These 2 prototypes were used to further explain the technology and platform concepts. The static screenshot-based prototype showed prospective categories (eg, identity, residency, medical profile, employment history, education, and children/preferred contacts) for the different types of documents. We also showed screenshots of a request from a provider for permission to see a user's data, accepting and giving the provider permission to view specific documents, and a transaction history showing that a document was shared with a provider.

At this stage in our project, the high-fidelity prototype platform was then used to demonstrate uploading and viewing a picture of a driver's license via the platform. The prototype platform for refugees had not been modified at all for our use case. Thus, our demonstration was performed in a sandbox or in a closed, nonlive testing environment to safely experiment with Web or software projects. We additionally decided to conduct further tests with sample documents until the technology was fully configured and free of errors.

Understanding and acceptance of metrics were measured with participation consent forms and qualitative feedback through open-ended questions.

Figure 2. Distributed ledger technology prototype.

Testing Assumptions 3, 6, and 7: Stakeholder Participation and Organizational/Data Standards

We tested assumptions 3, 6, and 7 regarding stakeholder participation and organizational and data standards through interview questions with representatives of the Downtown Austin Community Court, Central Health, and ECHO who interact with and provide services to PEH regularly (Multimedia Appendix 1). The participants were initially explained the technology platform, its relevant features, and our assumptions to test its practicality in addressing PEH identity management for health care and social services. Each representative was provided an opportunity to respond to the semistructured, facilitated discussion.

Results

Test of Assumptions

As was clear in the Methods section, multiple assumptions were tested in each of the approaches. As such, we note which assumption(s) was being tested, where appropriate. In particular, note that assumption 6 about data availability facilitating service transactions permeates most aspects of the testing process. The assumptions are listed in Textbox 1, and more detailed testing results are listed in Multimedia Appendix 2.

Assumptions 1 and 5

In testing assumptions 1 and 5, both of our homeless participants were able to set up a user account, accept a connection request, deny and accept a provider connection request, and exchange transactions to and from providers. Specific observations, difficulties encountered, and their relevance to our project and testing assumptions are displayed in Multimedia Appendices 2 and 3. During the second session, with modifications to the test, some of the onboarding workflow resulting from entering extensive personal information created a more streamlined user experience in terms of finding and categorizing types of documents. However, testers still experienced several challenges (Multimedia Appendices 2 and 3) in addition to those previously documented in Multimedia Appendix 2 regarding platform features and functionality. In general, the problems that users experience tend to fall into 5 buckets: trusting self-uploaded documents, understanding blockchain concepts, platform performance issues, confusion about the workflow, and privacy concerns. We return to this in the discussion below.

Overall, users on the platform were able to accomplish the tasks expected of them even if guidance was required, but there were definite issues that need to be solved. Trying to use an existing platform developed to assist refugees without modification did not translate to the use case of working with those experiencing homelessness. Furthermore, privacy and security did indeed turn out to be of paramount importance to end users.

Assumptions 2 and 6

The results of our tests showed that the homeless can understand the platform and its potential benefits, and a majority of individuals consented to participate immediately or at a later date. Feedback from our PEH test participants included the need for a wider variety of documents that can be validated and authenticated on the platform, leading us to consider possibilities for future expansion, such as involving state or federal institutions. We also planned to create benchmark documents facilitating services such as housing, disability benefits, Supplemental Nutrition Assistance Program (SNAP), medical, court documents, and insurance. Each benchmark document would consist of validation from multiple providers that the documents necessary for a service have been obtained and authenticated, increasing efficiency for service providers and improving access to services for their homeless clients.

Assumptions 3 and 6

The results showed that providers will accept data shared through the blockchain platform, but permissible document fidelity varies across organizations. Specifically, the requirements of what would be considered a certified document varied (see Table 3). Transactions requiring hard copies included employment documents, bank transactions, social security, and vital records, which would be difficult to include without altering the standards for data. The Medical Access Program Card (a local health insurance card), SNAP application documents, and housing applications permit document copies and can easily facilitate transactions without adjusting data standards.

The need for interorganizational records or documents from clients also varied. Many providers kept internal digital copies of client records and provided services to the homeless without proof of identity because of flexible policies for this population. Identification documents are unnecessary for social security, disability benefits, transitional housing, or rehabilitation services, but the lack of identification documents can delay medical care at hospitals. These providers envisioned a role for a distributed ledger system when clients did not grant permission for document storage and for time-sensitive documents such as a current utility bill needed to prove residency. In these cases, system-level efficiency could be improved if provider employees knew what type of documents were available and if documents of interest could easily be located without looking through extraneous information.

Table 3. Level of document certification

Certification level	Definition	Requirements to meet level of certification	Examples
Validated	Approved to be valid document and belong to the holder (platform user)	Documents coming directly (digitally and physically) from the originating Source	• Birth certificate uploaded and sent by Vital Records • Medical record sent by a doctor • Medical Assistance Program (MAP) card sent by the issuing entity
Verified	Verified that document belongs to the holder (platform user) and document is what it claims to be	Physical document must be seen by the certifying entity	• ID that is seen in person by Austin Police Department (APD) and uploaded or sent to a user
Uncertified	Exists as a document/asset, but not marked as validated or verified	None	• A self-uploaded form, such as a Homelessness Statement

Assumption 4

Potential changes in the organizational workflow to accommodate a new system, such as our platform, was another area of friction. The HMIS provided by the US Department of Housing and Urban Development (HUD) is already widely used across the country as a centralized database to confidentially aggregate data on the homeless and housing provisions provided to these individuals. Copies of documents and IDs can be stored via HMIS, and organizations funded by HUD are required to use HMIS for client data. From the service provider perspective, blockchain is most useful for documents not currently stored in the HMIS database or for client personal use to keep track of future appointments and pending documents needed for services. This showed that the blockchain platform may, for the time being, work complementary to other existing databases or information systems, adding new features that are difficult to establish using legacy systems.

Assumptions 5 and 6

Technological concerns included the privacy and confidentiality of indelible and immutable client data for a service that might not be successful in the long term. Providers agreed that the technology showed value for the nonhomeless population as well, but opinions on whether the technological barriers were surmountable for the homeless were mixed. The most vulnerable members of the homeless population struggle to keep a phone and remember their email IDs and passwords, complicating platform access and use. There will always be a level of tension between ease of use and privacy/security concerns. However, adequately addressing the latter may help facilitate more practical solutions for the former. The readiness to overcome any specific requirements for the use of the solution by PEH seemed to be determined by how many of their problems were being solved effectively by using such a platform.

Assumption 7

Analysis of interview feedback elucidated the requirement of clearly defined roles for each organization. Not all providers wanted responsibility or additional work processes of account creation, document validation, and resolution of transactional or account legitimacy. Legal issues and worries regarding Health Insurance Portability and Accountability Act (HIPAA) compliance to share personal health information and other client data between organizations were further constraints on stakeholder roles, participation, and interorganizational efficiency. The questions around governance and standardized processes for onboarding and managing roles remained unresolved. It was not clear whether the City of Austin had the capacity to manage a blockchain platform or if there was another organization in the city, which had the capacity and trust of all other stakeholders and general citizens, including the PEH.

Assumptions 1 and 4

The results of our tests demonstrated that the platform we had chosen to use lacked core functionality and configuration specifically needed for our particular use case. The user experience was time consuming and not intuitive for tasks such as account creation and sharing of data, which will likely hinder client and provider understanding, acceptance, and adoption of the platform. Validated data could not be securely shared while protecting privacy and confidentiality; thus, data transactions on the blockchain platform cannot facilitate services or increase interorganizational efficiency. The nature of sharing and transacting data on this platform was specific to the original use case of small farmers and refugees and was not suitable for our purpose of securely sharing and storing documentation. As our contract with the platform provider was constrained to limited modifications because of the time frame and budget of the project, we were unable to make the necessary changes to continue testing this platform for our use case. Given these findings, testing assumption 4 on the equivalence of the 2 platform use cases was invalidated.

Discussion

Principal Findings

Secure, fast, and reliable sharing of validated health and identity data is crucial to improve the quality of life and health outcomes of the general population, particularly vulnerable populations such as those experiencing homelessness. Our project tested whether a blockchain-based platform had the functionality to manage permissioned access and distribution of data while empowering patients with control over their own records. This study also showed the need to address challenges in establishing and operationalizing a blockchain system before trial or full-scale production.

Although broader and timely access to health and identity records can be achieved through blockchain, there are costs in transferring to a new system and training professionals and patients on the best methods of use to improve efficiency and outcomes. In the underserved populations, initial adoption depends on a user-friendly interface and an end user experience accounting for varying levels of technological access and ability. Unfamiliarity with blockchain technology and usage also creates challenges in adoption for service providers. Understanding the basic principles of blockchain technology is necessary to confer trust in the system and allow changes in workflow, promoting organizational efficiency and preventing the burden of additional verification of authorized documentation. Institutions participating in a blockchain system must also mutually agree upon the size of data that may be stored or transferred on the blockchain to maximize system performance [7,12].

Our research showed that a blockchain can be used to manage personal and health data by facilitating interoperability, patient control of documents, and a record of consented document access while maintaining data privacy and security. A validated, immutable, and decentralized ledger promotes system and transactional trustworthiness but cannot guarantee the absence of falsifications or errors from the point of origin. These mistakes are perpetuated in the blockchain without a manual content verification procedure [10,22,23]. Our results show that a private, permission-based blockchain might be suitable for sensitive personal and health information with regulatory guidelines and standards to ensure appropriate use of data. In a permissioned system, participating organizations need to decide who is responsible for the creation of new accounts [10]. Individual users may also request for their data to be erased [11]. If the data are protected health information (PHI), HIPAA mandates its deletion in the event of unauthorized access. PHI must also be destroyed when a data storage device is decommissioned. This is only possible when documents are not stored in the blockchain. If data are stored in an off-blockchain data repository or database, a record of the existence of deleted data may still be maintained within the chain [24]. There is ambiguity regarding whether metadata of PHI are considered PHI [25], but legal counsel with respect to the application of HIPAA and data privacy standards is vital to ensure compliance with regulatory frameworks [24,26]. A potential limitation regarding these data and use standards arises with respect to a private, partially centralized blockchain. To achieve compliance and vendor neutrality, an outside enforcing authority may be required [12].

Limitations

Timeline and budgetary constraints from the Mayor’s Challenge Project limited our ability to fully modify the platform for our use case. We were unable to develop an end user experience promoting homeless client adoption or develop changes in workflow, policy, and data sharing agreements for the best use of blockchain technology. The significant security and usability issues with our original platform prevented large-scale implementation in an empirical setting.

Further research is required to compare the efficacy and costs of an approach based on blockchain technology with other alternative approaches. Although the research presented here shows the potential of a blockchain-based approach, we need to better understand the comparative benefits and costs. For example, the City of Austin could have created a central database to which all partners and collaborators would agree to add their data. Questions on whether all participants would adopt the use of standard data types and a centralized, city-owned database would need to be addressed. Experience has shown that not all advocacy groups or people experiencing homelessness fully trust the city government, for example. The purpose here, however, was to show whether a blockchain-based approach could work, and it was not to address the question of comparative effectiveness.

Comparison With Prior Work

Several studies to date have explored the potential applications of blockchain to solve key issues in the health care sector. This study demonstrates a methodology and rigor that may be needed to test if a blockchain can be used to securely store and track verified documentation, promote client ownership of data, and improve interoperability by facilitating permission-based data sharing. This study paves the way for future studies by detailing specific organizational, logistical, and system considerations for successful and scalable implementation.

Conclusions

Blockchain may provide a means for consented access to validated personal and health data, thus increasing interoperability without compromising the security or privacy of data [12,24]. Existing solutions need to be put through rigorous testing before being adopted at scale. We developed assumptions based on feedback from end users, PEH, and service providers on important aspects to be tested. We engaged these groups actively to test the assumptions using a preexisting platform and found that many of the assumptions could not be validated, given the constraints of the platform, limitations on time and resources in the pilot, and lack of clarity on legal and compliance implications of this new technology. On the basis of our learnings through this pilot study, we opine that using an off-blockchain data lake or extant provider databases for PHI storage and systematic storage of an index of health records and associated metadata on the chain can permit the management of access and data control while complying with privacy and regulatory standards [24,27]. Maintaining limited personal data on the blockchain maximizes the speed of transactions and scalability of the blockchain system [11,12,20]. Through an API, organizations can integrate data and receive accurate, updated information in a usable format [19]. Faster availability of real-time data reduces delays in service [10,11,28] and promotes coordinated health care and specialized treatment based on outcomes and efficacy. Shifting data ownership and control to the individual optimizes access to health and social services and engages a patient in their own care through selective sharing of information and data with providers or researchers [29-31]. Patient-reported measurable outcomes and data from mobile apps or on-person sensors may also be integrated, creating a single access point for all real-time health data and improving personalized health care [12,13]. These benefits outweigh the challenges in adoption, employment, and investment of a blockchain system. The application of current recommendations and continued research into blockchain implementation is crucial to develop cost-effective strategies for the operationalization of blockchains while ensuring efficiency, data privacy, and scalability in the health care ecosystem [24].

Acknowledgments

The authors would like to thank Bloomberg Philanthropies for their financial support through the Mayor’s Challenge Program, City of Austin’s Office of Technology Innovation, Austin Mayor’s Office, Austin/Travis County Emergency Medical Services, Central Health, and PEH who participated in the study.

Authors' Contributions

AK and VR designed the study along with City of Austin partners. AK, SA, and VR wrote the manuscript. AK, SA, and VR reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview questions for service providers.
[PNG File , 277 KB - jmir_v22i6e16887_app1.png]

Multimedia Appendix 2
Platform testing session results.
[PNG File , 431 KB - jmir_v22i6e16887_app2.png]

Multimedia Appendix 3
Second session findings.
[PNG File , 322 KB - jmir_v22i6e16887_app3.png]

References

1. Ending Community Homelessness Coalition. 2016 May 17. Ending Community Homelessness Coalition 2016 Annual Report URL: http://www.webcitation.org/786921PpK [accessed 2019-05-03] [WebCite Cache ID 786921PpK]
2. Khurshid A, Gadnis A. Using blockchain to create transaction identity for persons experiencing homelessness in America: policy proposal. JMIR Res Protoc 2019 Mar 6;8(3):e10654 [FREE Full text] [doi: 10.2196/10654] [Medline: 30839279]
3. Angraal S, Krumholz HM, Schulz WL. Blockchain technology: applications in health care. Circ Cardiovasc Qual Outcomes 2017 Sep;10(9):e003800. [doi: 10.1161/CIRCOUTCOMES.117.003800] [Medline: 28912202]
4. Funk E, Riddell J, Ankel F, Cabrera D. Blockchain technology: a data framework to improve validity, trust, and accountability of information exchange in health professions education. Acad Med 2018 Dec;93(12):1791-1794. [doi: 10.1097/ACM.0000000000002326] [Medline: 29901658]
5. Kamel Boulos MN, Wilson JT, Clauson KA. Geospatial blockchain: promises, challenges, and scenarios in health and healthcare. Int J Health Geogr 2018 Jul 5;17(1):25 [FREE Full text] [doi: 10.1186/s12942-018-0144-x] [Medline: 29973196]
6. Košťál K, Helebrandt P, Belluš M, Ries M, Kotuliak I. Management and monitoring of IoT devices using blockchain. Sensors (Basel) 2019 Feb 19;19(4):856 [FREE Full text] [doi: 10.3390/s19040856] [Medline: 30791392]
7. Krawiec R, White M, Filipova M, Quarre F, Barr D. Deloitte. 2016. Blockchain: Opportunities for Health Care URL: https://www2.deloitte.com/us/en/pages/public-sector/articles/blockchain-opportunities-for-health-care.html [accessed 2019-05-06] [WebCite Cache ID 78AaeFsla]
8. Mertz L. (Block) chain reaction: a blockchain revolution sweeps into health care, offering the possibility for a much-needed data solution. IEEE Pulse 2018;9(3):4-7. [doi: 10.1109/MPUL.2018.2814879] [Medline: 29757744]
9. Roman-Belmonte JM, de la Corte-Rodriguez H, Rodriguez-Merchan EC. How blockchain technology can change medicine. Postgrad Med 2018 May;130(4):420-427. [doi: 10.1080/00325481.2018.1472996] [Medline: 29727247]
10. Zhuang Y, Sheets L, Shae Z, Tsai JJ, Shyu C. Applying blockchain technology for health information exchange and persistent monitoring for clinical trials. AMIA Annu Symp Proc 2018;2018:1167-1175 [FREE Full text] [Medline: 30815159]

11. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. J Med Internet Res 2019 Feb 8;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
12. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing blockchains for efficient health care: systematic review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
13. Dimitrov DV. Blockchain applications for healthcare data management. Healthc Inform Res 2019 Jan;25(1):51-56 [FREE Full text] [doi: 10.4258/hir.2019.25.1.51] [Medline: 30788182]
14. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]
15. Hylock RH, Zeng X. A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study. J Med Internet Res 2019 Aug 31;21(8):e13592-e13595 [FREE Full text] [doi: 10.2196/13592] [Medline: 31471959]
16. Dexter S. Mango Research-Blockchain Explained Simply!. 2018. How Are Blockchain Transactions Validated? Consensus vs Validation URL: https://www.mangoresearch.co/blockchain-consensus-vs-validation/ [accessed 2019-05-03] [WebCite Cache ID 786Ae8d7C]
17. Pérez-Solà C, Delgado-Segura S, Navarro-Arribas G, Herrera-Joancomartí J. Another coin bites the dust: an analysis of dust in UTXO-based cryptocurrencies. R Soc Open Sci 2019 Jan;6(1):180817 [FREE Full text] [doi: 10.1098/rsos.180817] [Medline: 30800345]
18. Wood G. Semantic Scholar. 2014. Ethereum: A Secure Decentralised Generalised Transaction Ledger URL: https://pdfs.semanticscholar.org/259c/790cfa0b7a73836812217ad149d34aa09aa6.pdf?_ga=2.191984810.127702914.1586246225-1679671381.1567599385 [accessed 2019-05-06] [WebCite Cache ID 78AcHc1hD]
19. O'Dowd E. HITInfrastructure. 2016. Why Application Programming Interfaces Are Key for Healthcare URL: https://hitinfrastructure.com/features/why-application-programming-interfaces-are-key-for-healthcare [accessed 2019-05-03] [WebCite Cache ID 786A5nuc8]
20. Gordon WJ, Catalini C. Blockchain technology foBlockchain technology for healthcare: facilitating the transition to patient-driven interoperabilityr healthcare: facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
21. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 1;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
22. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. How Blockchain Could Empower eHealth: An Application for Radiation Oncology. In: Proceedings of the VLDB Workshop on Data Management and Analytics for Medicine and Healthcare. 2017 Presented at: DMAH'17; September 1, 2017; Munich, Germany p. 3-6. [doi: 10.1007/978-3-319-67186-4_1]
23. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
24. Meinert E, Alturkistani A, Foley KA, Osama T, Car J, Majeed A, et al. Blockchain implementation in health care: protocol for a systematic review. JMIR Res Protoc 2019 Feb 8;8(2):e10994 [FREE Full text] [doi: 10.2196/10994] [Medline: 30735146]
25. Esposito C, de Santis A, Tortora G, Chang H, Choo KR. Blockchain: a panacea for healthcare cloud-based data security and privacy? IEEE Cloud Comput 2018 Jan;5(1):31-37. [doi: 10.1109/MCC.2018.011791712]
26. Murphy K. EHRIntelligence. 2016. How Health IT Standards Enable Patient Access to Health Data URL: https://ehrintelligence.com/news/how-health-it-standards-enable-patient-access-to-health-data [accessed 2019-05-03] [WebCite Cache ID 786AGKJgE]
27. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski A, Botezatu A, Prikhodko P, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget 2018 Jan 19;9(5):5665-5690 [FREE Full text] [doi: 10.18632/oncotarget.22345] [Medline: 29464026]
28. de Salvo K. HealthIT. 2015. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap URL: https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf [accessed 2019-03-05] [WebCite Cache ID 78AeSxDNr]
29. Engelhardt MA. Hitching healthcare to the chain: an introduction to blockchain technology in the healthcare sector. TIM Rev 2017 Oct 27;7(10):22-34. [doi: 10.22215/timreview/1111]
30. Kitson A, Marshall A, Bassett K, Zeitz K. What are the core elements of patient-centred care? A narrative review and synthesis of the literature from health policy, medicine and nursing. J Adv Nurs 2013 Jan;69(1):4-15. [doi: 10.1111/j.1365-2648.2012.06064.x] [Medline: 22709336]
31. Prajna SS. Comms MEA. 2017. Unleashing the Power of the Ledger URL: https://www.commsmea.com/16927-unleashing-the-power-of-the-ledger [accessed 2019-05-06] [WebCite Cache ID 78Aem4zko]

Abbreviations

API: application programming interface

DLT: distributed ledger technology
ECHO: Ending Community Homelessness Coalition
HIPAA: Health Insurance Portability and Accountability Act
HMIS: Homeless Management Information System
HUD: US Department of Housing and Urban Development
PEH: persons experiencing homelessness
PHI: protected health information
SNAP: Supplemental Nutrition Assistance Program

Edited by G Eysenbach; submitted 03.11.19; peer-reviewed by W Zhang, TT Kuo; comments to author 23.12.19; revised version received 25.02.20; accepted 26.02.20; published 04.06.20.

Please cite as:
Khurshid A, Rajeswaren V, Andrews S
Using Blockchain Technology to Mitigate Challenges in Service Access for the Homeless and Data Exchange Between Providers: Qualitative Study
J Med Internet Res 2020;22(6):e16887
URL: https://www.jmir.org/2020/6/e16887
doi:10.2196/16887
PMID:32348278

Original Paper

An Architecture and Management Platform for Blockchain-Based Personal Health Record Exchange: Development and Usability Study

Hsiu-An Lee[1,2,3,4,5], MS; Hsin-Hua Kung[2,3,4,5], BS; Jai Ganesh Udayasankaran[3,4,6], MSc, MBA; Boonchai Kijsanayotin[3,4,7], MSc, MD, PhD; Alvin B Marcelo[3,4,8], MD; Louis R Chao[1], PhD; Chien-Yeh Hsu[2,3,4,5,9], PhD

[1]Department of Computer Science and Information Engineering, Tamkang University, New Taipei City, Taiwan
[2]Taiwan e-Health Association, Taipei, Taiwan
[3]Asia eHealth Information Network, Hong Kong, Hong Kong
[4]Standards and Interoperability Lab, Smart Healthcare Center of Excellence, Taipei, Taiwan
[5]Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan
[6]Sri Sathya Sai Central Trust, Prasanthi Nilayam, Puttaparthi, India
[7]Thai Health Information Standards Development Center, Health System Research Institute, Ministry of Public Health, Bangkok, Thailand
[8]University of the Philippines, Manila, Philippines
[9]Taipei Medical University Master Program in Global Health and Development, Taipei, Taiwan

Corresponding Author:
Chien-Yeh Hsu, PhD
Department of Information Management
National Taipei University of Nursing and Health Sciences
No 365, Ming-te Road, Peitou District, Taipei City
Taipei, 112
Taiwan
Phone: 886 939193212
Email: cyhsu@ntunhs.edu.tw

Abstract

Background: Personal health record (PHR) security, correctness, and protection are essential for health and medical services. Blockchain architecture can provide efficient data retrieval and security requirements. Exchangeable PHRs and the self-management of patient health can offer many benefits to traditional medical services by allowing people to manage their own health records for disease prevention, prediction, and control while reducing resource burdens on the health care infrastructure and improving population health and quality of life.

Objective: This study aimed to build a blockchain-based architecture for an international health record exchange platform to ensure health record confidentiality, integrity, and availability for health management and used Health Level 7 Fast Healthcare Interoperability Resource international standards as the data format that could allow international, cross-institutional, and patient/doctor exchanges of PHRs.

Methods: The PHR architecture in this study comprised 2 main components. The first component was the PHR management platform, on which users could upload PHRs, view their record content, authorize PHR exchanges with doctors or other medical health care providers, and check their block information. When a PHR was uploaded, the hash value of the PHR would be calculated by the SHA-256 algorithm and the PHR would be encrypted by the Rivest-Shamir-Adleman encryption mechanism before being transferred to a secure database. The second component was the blockchain exchange architecture, which was based on Ethereum to create a private chain. Proof of authority, which delivers transactions through a consensus mechanism based on identity, was used for consensus. The hash value was calculated based on the previous hash value, block content, and timestamp by a hash function.

Results: The PHR blockchain architecture constructed in this study is an effective method for the management and utilization of PHRs. The platform has been deployed in Southeast Asian countries via the Asia eHealth Information Network (AeHIN) and has become the first PHR management platform for cross-region medical data exchange.

Conclusions: Some systems have shown that blockchain technology has great potential for electronic health record applications. This study combined different types of data storage modes to effectively solve the problems of PHR data security, storage, and transmission and proposed a hybrid blockchain and data security approach to enable effective international PHR exchange. By partnering with the AeHIN and making use of the network's regional reach and expert pool, the platform could be deployed and promoted successfully. In the future, the PHR platform could be utilized for the purpose of precision and individual medicine in a cross-country manner because of the platform's provision of a secure and efficient PHR sharing and management architecture, making it a reasonable base for future data collection sources and the data analytics needed for precision medicine.

(J Med Internet Res 2020;22(6):e16748) doi:10.2196/16748

KEYWORDS

blockchain; personal health records; health information interoperability; precision health care; health information management

Introduction

Background

Traditionally, standard clinics have offered medical services focused on disease treatment. However, with the world's current aging populations, there is a growing gap between what services clinics offer and patients' actual needs. This means that clinics may not be equipped to offer the complete range of care required by patients, resulting in preventable medical harm. The National Institute for Health and Care Excellence's 2016 Multimorbidity Clinical Assessment and Management Guidelines Report [1] emphasized the importance of integrating patient-centered decision-making methods for multiple problems, with a focus on precision medicine. Precision medicine is a disease treatment and prevention strategy formulated with reference to individual variability in terms of genes, environment, and lifestyle, which is used to determine necessary dynamic changes and personalized treatment for preventative health care and clinical care. The core elements of precision medicine are historical disease data, daily vital signs data, personal health management, and medical record exchange, and it aims to stop potentially harmful or unnecessary medical behavior, integrate care, reduce treatment burden, and help patients select meaningful treatment and care goals through accurate assessment. With the requirements of precision medicine mentioned earlier, there is a need to not only maintain patients' electronic medical records (EMRs) in hospitals but also to establish personal health records (PHRs) by combining medical records from different health institutes and functions of precision medicine, which patients can use to save, manage, use, and exchange with health care practitioners.

PHRs are highly private data, and this sensitivity means that there are significant security challenges involved in their management and exchange. Any system that seeks to manage and exchange such records must ensure that health records are exchanged appropriately, that they are not leaked, and that protected data are not tampered with. A good way to achieve the secure exchange of health records is by using blockchain architecture. A decentralized storage management architecture based on blockchain would be able to meet the security requirements. In a 2016 study, Ford [2] predicted that 75% of the adults worldwide will be using PHRs by 2020 without any external incentives. The importance of a PHR is that it allows a health care provider to examine a patient's history of illnesses and medications and it provides a basis for medical decision making. More importantly, PHRs offer a basis for personal health management. PHRs include various health information such as medical information, vital signs (heartbeat, blood pressure, blood sugar, and body temperature), family disease history, and blood test reports [3-5]. Most countries today, however, still use the EMR system. In 2013 in Taiwan, a total of 502 hospitals had a comprehensive EMR system for accessing medical records, inspection reports, medical images, medication information, and so on. However, these data only exist in hospitals and are exchanged between other hospitals or clinics via the EMR exchange center. To achieve the goals of precision medicine and health care, a *patient-centered* approach to record management and exchanges is required; the traditional centralized PHR repository in hospitals does not meet the requirements to achieve this. A patient-centered approach would involve PHRs being managed by the patients themselves, while providing those records to various health care providers as needed. This kind of system would require a very secure architecture to protect PHR data.

According to the National Health Insurance (NHI) Administration of the Ministry of Health and Welfare in Taiwan, the average number of outpatient visits, not including Chinese medicine or dentists, is 13 per year for people in Taiwan. Most of these people visit different hospitals for treatment of the same condition over a short period of time. With the PHR system, people can manage their own health records and conditions, and doctors can also view their past medical records and medication status.

Blockchain technology was proposed by Nakamoto in 2008 [6] in a white paper titled "Bitcoin: A Peer-to-Peer Electronic Cash System." A blockchain has the characteristics of decentralization, and its encryption mechanism can be designed to verify the data content to ensure that the data have not been tampered with. In this paper, the blockchain concept was used to solve the problem of data security and third-party authentication in the transaction process. A blockchain is a decentralized public account that records all money transactions and how much money everyone owns. John et al [7] proposed that the use of blockchain technology in electronic health care records can avoid the need to add another organization between the patient and the records. It is not a new repository for data but rather implies a decentralized control mechanism in which all users have an interest, but no one exclusively owns the data. This technology can improve data safety and remove privacy issues. Pouyan et al [8] stated that regarding the trust in health information exchange competency and exchange integrity, the blockchain architecture is more trustworthy than other exchange mechanisms for exchanging highly sensitive information.

This design differed from previous work on blockchain infrastructures and associated consensus mechanisms in that, while they operate in a decoupled manner from other blockchain frameworks, Fast Healthcare Interoperability Resource (FHIR) Chain [9] focuses on designing the decisions of smart contracts to be compatible with any existing blockchain architecture that supports the execution of smart contracts. However, this architecture remains vulnerable to the 51% of cyber attacks and does not provide complete data security.

Objectives

This study proposed a blockchain-based architecture for storing, sharing, and protecting sensitive personal information. In the proposed architecture, the blockchain manages the authorization of data exchanges between patients, health care providers, and other users. The blockchain does not physically replace the electronic health record system, as most hospital information systems store detailed EMRs in a secure database on site or on a duplicate site located outside the hospital. Therefore, the blockchain architecture simply helps to ensure the security, confidentiality, integrity, and availability of the data. Combined with FHIR's data format standards, stakeholders can read and write data into their own electronic health record systems that can be exchanged securely with other systems using the blockchain. The computational strength of the encryption built into the blockchain ensures that the data are correctly and safely transferred during PHR exchange transactions. However, a blockchain is not a data repository, rather it is a ledger of data integrity. This technology can be used to exchange records, verify data, and protect sensitive data. It can ensure that medical records will not be modified by unauthorized third parties. The uploading time of the data to the blockchain can also be recorded. Thus, the enabling of the collection of a patient's more complete longitudinal data and the ability to share it remotely with professionals can allow for better decision making and reduce medical errors and medical malpractice.

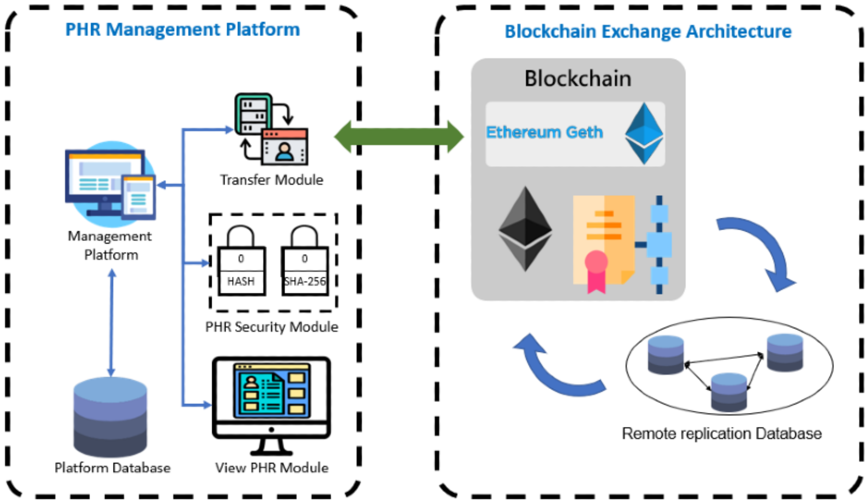
Methods

The blockchain-based exchange architecture for PHR management proposed in this study comprises 2 main components. The first component is the PHR management platform, on which users can upload PHRs, view their record content, authorize PHR exchange with doctors or other medical health care providers, and check their block information. When a PHR is uploaded, the hash value of the PHR is calculated by the SHA-256 algorithm, and the PHR is encrypted by the RSA (Rivest-Shamir-Adleman) encryption mechanism before being transferred to a secure database. The second component of the architecture is the blockchain exchange architecture, which is based on Ethereum to create a private chain. Proof of authority (PoA), which delivers transactions through a consensus mechanism based on identity, is used for consensus. The hash value is calculated based on the previous hash value, block content, and timestamp by a hash function.

The architecture of the platform is shown in Figure 1. The PHR management platform consists of the transfer module, the security module, and the view PHR module. The transfer module allows users to connect to the blockchain exchange architecture to create or search for blocks. The security module is used to encrypt and confirm the PHR content. The view PHR module displays the PHR content for personal health management or for doctors to view the record.

The blockchain architecture in this study is designed based on Ethereum, including elliptic curve digital signature, PoA, and the new block creation function. The blockchain architecture ensures that the PHR content remains secure and confirms that the PHR content is correct.

Figure 1. Personal health record management platform and blockchain architecture. PHR: personal health record.



Personal Health Record Management Platform

The major goal of this study was to build a cross-area health information exchange platform that could fulfill the needs of international medical services. This study used My Health Bank (MHB) as an initial example of PHRs. In Taiwan, MHB is issued by the NHI and contains a majority of the clinical data collected from different health care services. MHB not only includes the necessary clinical data chronically arranged by time for a single patient but also contains the information entered by the patient, such as blood pressure measured at home. Therefore, there was a good reason for this study to choose MHB for the PHRs in the Asia eHealth Information Network (AeHIN). Detailed items of the MHB are provided in this manuscript. Basically, PHRs refer to individual-centric personal health data from different medical service providers or devices, while EMRs represent the data of a patient in a single hospital.

Multiple simulated computers are used as blockchain nodes in this study to emulate the encryption and secure storage of a PHR in this study. As health records are private data, the blockchain must be built in a secure environment as a private chain, increasing the efficiency and stability of data transmission and sharing.

MHB was used as a PHR example in this study. MHB was launched by the Ministry of Health and Welfare of Taiwan in 2015. It allows Taiwan's NHI members to download their own health records from its website.

The MHB data contain all the necessary clinical information because they are generated by the hospital when it applies for health insurance payments.

The entire PHR of any single patient was uploaded in our platform. For authority management and confidentiality, we used a variety of tags in the contents to specify the function levels to different uses through a carefully designed user interface, through which patients could assign which data would not be revealed to others as well as assign tags to the data. Our design to keep the whole data is for the purpose of future use of the data, as the PHR platform could also become a clinical data repository and the data could be used for further analysis of precision medicine in the future.

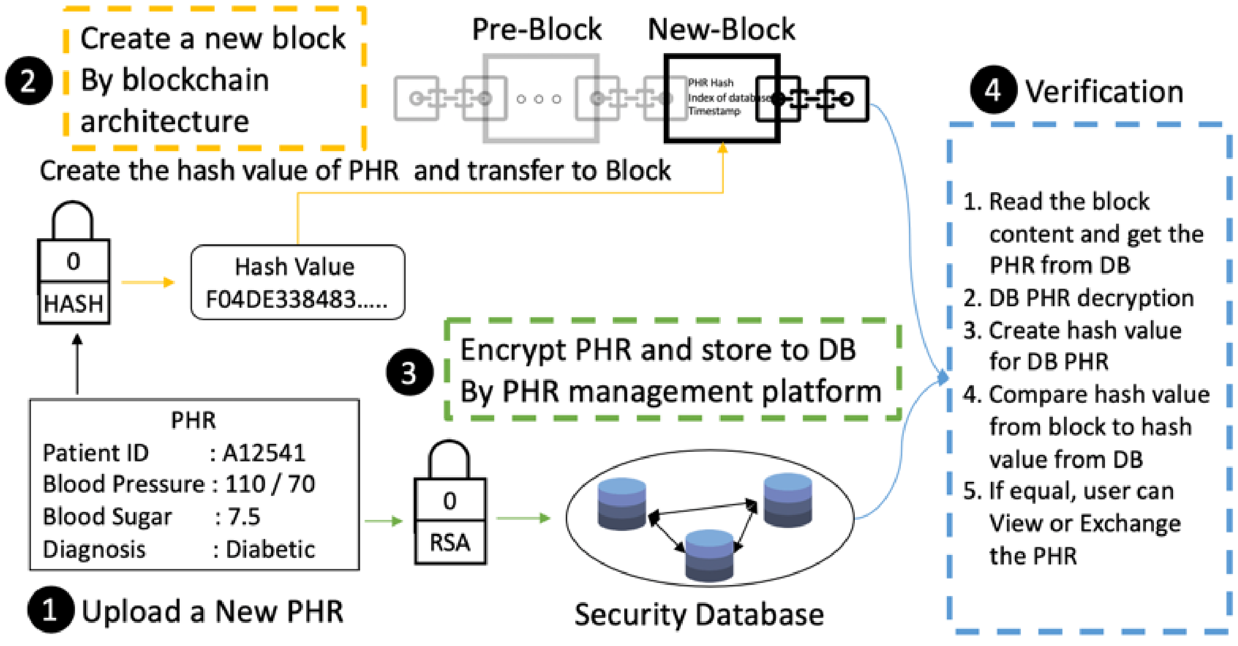
The MHB data include (1) outpatient information for Western medicine, traditional Chinese medicine, and dentistry; (2) hospitalization information; (3) allergy information; (4) images and information for pathological exams and tests; (5) patients' discharge record abstract; (6) patients' intention for organ donation and palliative care; (7) preventive health data; (8) preventive vaccination information; (9) patients' health insurance card information; (10) premium and charging specific information; and (11) insurance premium payment specific information. The MHB file format can be selected as either XML or JSON. This study used the XML format.

Hash Value for Data Integrity Confirmation

To ensure that PHRs are not modified when they are transferred between platforms, this study designed a hash function to confirm the integrity of PHR data. SHA-256 was used to create a hash value for each PHR. SHA-256 is a cryptographic hash function, which takes an input and produces a 256-bit (32-byte) hash value known as a message digest, typically rendered as a hexadecimal number, 40-digit long. It was designed by the United States National Security Agency and is a US Federal Information Processing Standard [10,11]. If the PHR data have not been altered during transfer, the SHA-256 hash value would remain the same. Unlike encryption, which converts text into reversible cipher texts of different lengths, the hash function converts text into irreversible hash strings (or message digests) of the same length.

When users upload their PHRs to the platform, the PHR hash value is created and transferred to the blockchain architecture as block content. Then, when the PHRs are viewed by the owner, or exchanged with other users, the platform obtains the hash value from the block and calculates the PHR hash value by SHA-256 again. If the hash value from the PHR is equal to the hash value from the block, the PHR data have not been modified. The procedure of PHR management is shown in Figure 2.

Figure 2. Personal health record creation, uploading, and verification procedure. DB: database; PHR: personal health record.



Data Encryption for Personal Health Record Security

In this study, PHRs were encrypted by RSA before being uploaded to the secure database. RSA is a public-key cryptosystem used for secure data transmission [12]. The encryption key is public and differs from the decryption key, which is private. The platform automatically creates the RSA public and private keys for users. When users upload their PHRs, the public key is used to encrypt the record. Thus, even if a malicious attacker were to overcome the firewall and all other security mechanisms, they would only be able to obtain the encrypted PHR and would have no means of decrypting it. The user private key is used to decrypt the PHR when exchanged.

Viewing Personal Health Record and Block Information for Personal Health Management

This study designed a PHR exchange architecture in which PHR contents are not read when users upload their PHRs; the platform only uploads encrypted PHRs to the secure database, thus ensuring the security of personal data.

Moreover, this study developed a user interface for personal health management that shows PHR contents when users want to access them. Using MHB as an example, when users use the application to read their PHRs, it means that the platform has the authority to read the PHRs. The PHRs are then decrypted by the user's RSA private key, and the platform reads PHR data, without storing them. This means that the platform cannot simply access PHRs without explicit user consent and action.

Blockchain Exchange Architecture

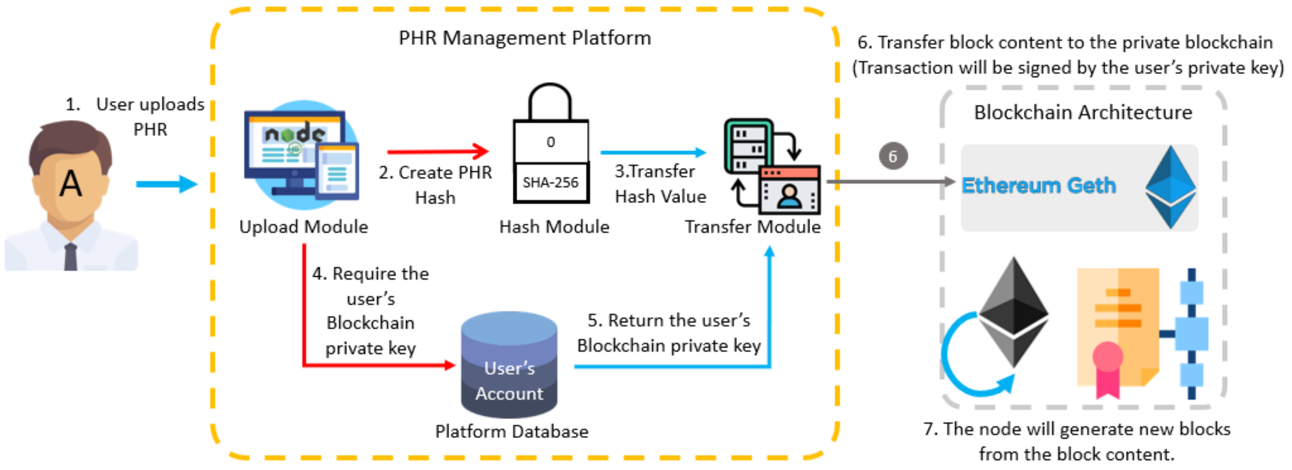
As the blocks in a blockchain cannot be tampered with or maliciously altered, this study stored PHR hash values in a blockchain to protect the PHR data and confirm the integrity of the PHR contents. Ethereum's private chain was used as the blockchain architecture, and the Geth (Go Ethereum) application, which is the Ethereum protocol, was used to transfer the transaction from the proposed platform to the blockchain exchange architecture, create a new block, and connect to the blockchain. The block creation process is shown in Figure 3.

To secure against private data being leaked during transmission on the network, the data are encrypted during the data transmission process. The health record uploaded to the secure database by the platform is also encrypted to ensure the privacy of the user. The block content includes the PHR hash and timestamp, where the PHR hash is used to check whether the PHR in the database has been tampered with. If a malicious attacker attempts to obtain the block content, they will only get a collection of random numbers. The encryption method combines hash encryption and asymmetric encryption. The block content is protected by a hash encryption function that uses SHA-256 to scramble data into a set of hexadecimal strings. Asymmetric encryption uses the elliptic curve digital signature algorithm to encrypt PHR transfer information, ensuring the integrity and nonrepudiation of transaction data, and then the PoA consensus mechanism is used for validation by a qualified verifier established by an audited authority to confirm the correctness and validity of the PHR and create the verified blocks of the blockchain.

Elliptic curve cryptography (ECC) is a public-key cryptography based on elliptic curve mathematics, also known as asymmetric cryptography. The elliptic curve digital signature algorithm is based on ECC for digital signatures. The working principle is similar to that of most digital signature algorithms. They are signed with a private key and verified with a public key, thus offering nonrepudiation. Compared with traditional digital signature algorithms (such as RSA), ECC is faster, offers stronger security, and requires shorter signatures.

In the proposed platform, each user has one password for a user account and a private key for blockchain and PHR decryption. To improve the platform efficiency, users can choose to store their personal blockchain private key in the platform's security database (or store it themselves). When data are uploaded to the platform, the system will retrieve the key from the database to complete the transaction process.

Figure 3. Block creation process. PHR: personal health record.



Proof of Authority for Block Creation

PoA is a technology that achieves consensus in a private chain. In the operation, an authorized node has the authority to generate the next block in a blockchain network. The blockchain information reaches the extreme value of the consensus of all nodes, which can guarantee that the latest blocks are accurately connected in series to the blockchain, and the blockchain information stored by the nodes is consistent, indivisible, and even resistant to malicious attacks. In this study, the private chain consensus mechanism is established, and the verifier is set on multiple simulated computers. The initial setup verifier node is set up on the simulated computers. In the feature, possible nodes represent cooperating institutions, medical institutions, research centers, and so on; the verifier uses this identity to obtain the right to verify.

Compared with other proof mechanisms, the key elements of the PoA network in this study include the following:

1. Improved efficiency: Block creation is accelerated and the waiting time for data exchange is reduced.
2. Verifier setup: A mutual supervision relationship with partner institutions is established to allow self-supervision and supervision of others, preventing the blockchain from being controlled by the node manager; the verifier can vote for a new verifier or remove an unqualified verifier at any time.
3. Highly scalable and highly compatible: It is also possible to complete intelligent collaborative construction and optimize it.

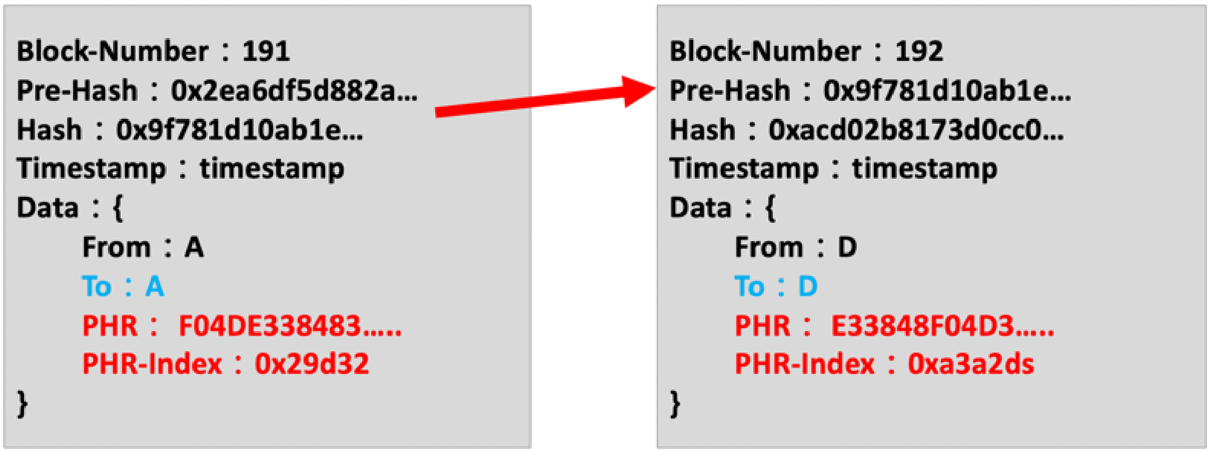
Hash Value for Block Corrected, Confirmed, and Connected

The cryptographic hash function is an important part of the blockchain. It is essentially a function that gives security capabilities to the created block, based on processed transactions, making them immutable. In Ethereum’s function, SHA-256 is used to create new blocks. The hash of a block is created based on the block content, previous hash value, and timestamp. The block content and architecture are shown in Figure 4.

Block content includes the following:

- Block number: Current block number
- Pre-Hash: The hash value of the previous block
- Hash: The hash value of this block
- Timestamp: Current time
- PHR hash: The hash value of PHR created by the platform
- PHR index: The index position of the health record in the secure database

Figure 4. Block content on the blockchain architecture. PHR: personal health record.



```
Block-Number : 191
Pre-Hash : 0x2ea6df5d882a…
Hash : 0x9f781d10ab1e…
Timestamp : timestamp
Data : {
      From : A
      To : A
      PHR : F04DE338483…..
      PHR-Index : 0x29d32
}
```

```
Block-Number : 192
Pre-Hash : 0x9f781d10ab1e…
Hash : 0xacd02b8173d0cc0…
Timestamp : timestamp
Data : {
      From : D
      To : D
      PHR : E33848F04D3…..
      PHR-Index : 0xa3a2ds
}
```

Overall System Workflow

Personal Health Record Exchange Authority Mechanism

Users can manage the authority for PHR exchange once they have uploaded their PHRs. When users want to make their PHRs available to a doctor, the authority assignment procedure is as shown in Figure 5.

The workflow of the system comprises 3 components: upload, exchange, and view. To begin, a user uploads their PHR to the platform (Figure 6).

In the uploading process, the PHR is assigned a hash value by SHA-256. Then, the PHR is transferred to the secure database after encryption by RSA. Once the data are stored in the database, blockchain is used to ensure data security and integrity. SHA-256 and ECC are then used to create a block, and the Ethereum architecture is used as the blockchain architecture in this study. The PHR hash value and the PHR index in the database are transmitted to the Ethereum block by the user’s blockchain account (public key) and using the user’s private key signature. To create a block, block content must be verified and the block hash value must be calculated by the verifier node; it is then broadcast to each node.

The workflow of users sharing their PHRs with a doctor is shown in Figure 6. First, the platform sends the transaction to the blockchain architecture. The block architecture will then select the user’s block and read its content. The users’ PHRs will be obtained from the secure database based on the database index of the PHRs and decrypted using the users’ private key, and the hash value will be created again. The PHR will then be transferred to the doctor after being encrypted by the doctor’s RSA public key and decrypted by the user’s private key if the hash value is equal to the block content’s hash.

The workflow of viewing the PHR content is shown in Figure 7. When users want to view their PHRs or share their PHRs with a doctor, the platform will send the transaction to the blockchain architecture. The blockchain architecture will confirm that the PHR content has not been modified and that the user or doctor has the authority to view the PHR. The PHR will then be transferred to the user or doctor after being encrypted by their RSA public key. The user or doctor will then use their own private key to decrypt it. They will then be able to view the PHR content. MHB is used as an example in this study.

Figure 5. Workflow of a user uploading their personal health record. PHR: personal health record.

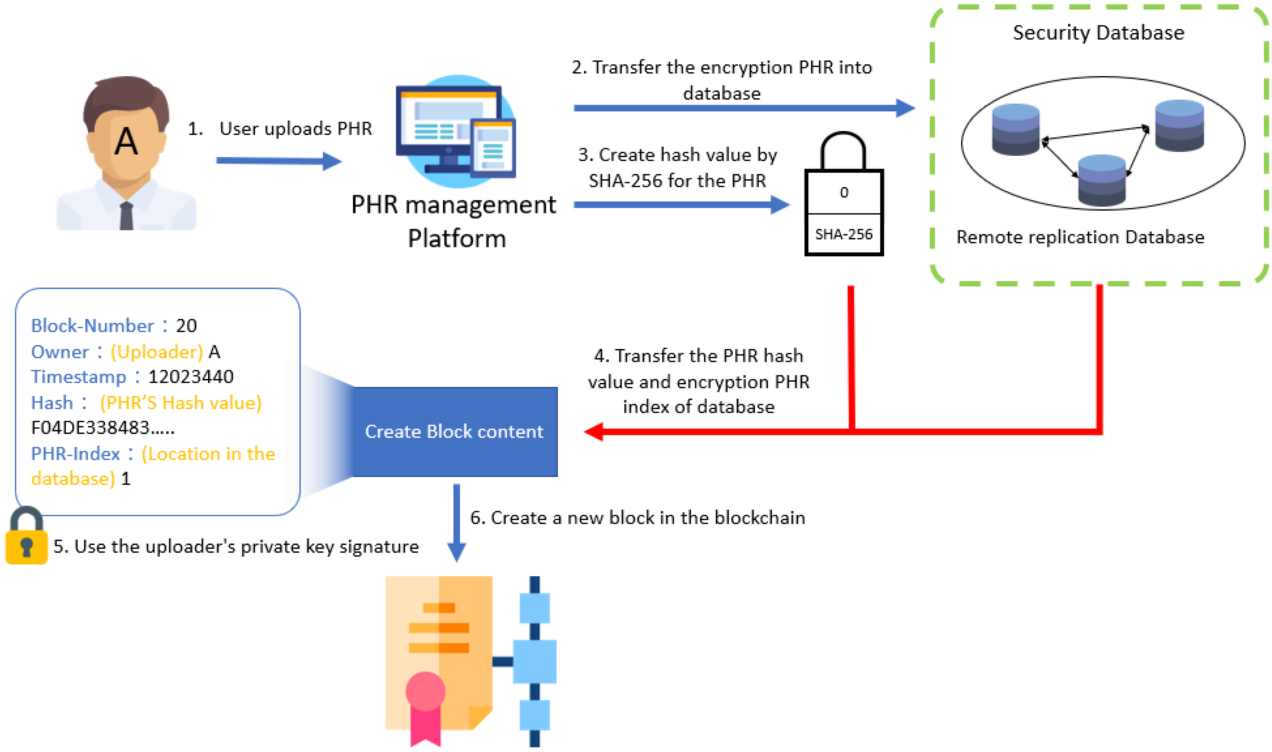


Figure 6. Workflow of a user sharing their personal health record with a doctor. PHR: personal health record; RSA: Rivest-Shamir-Adleman.

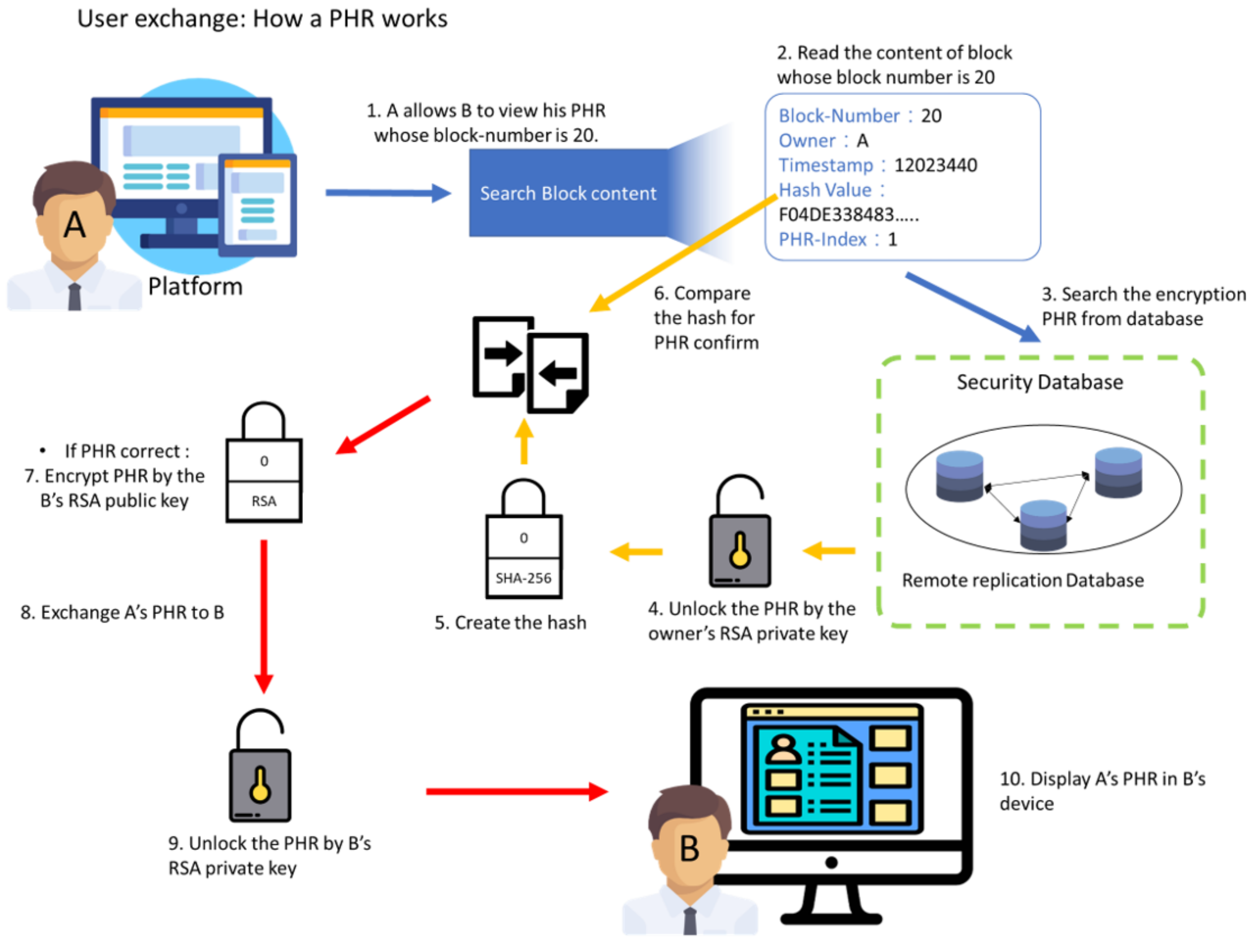
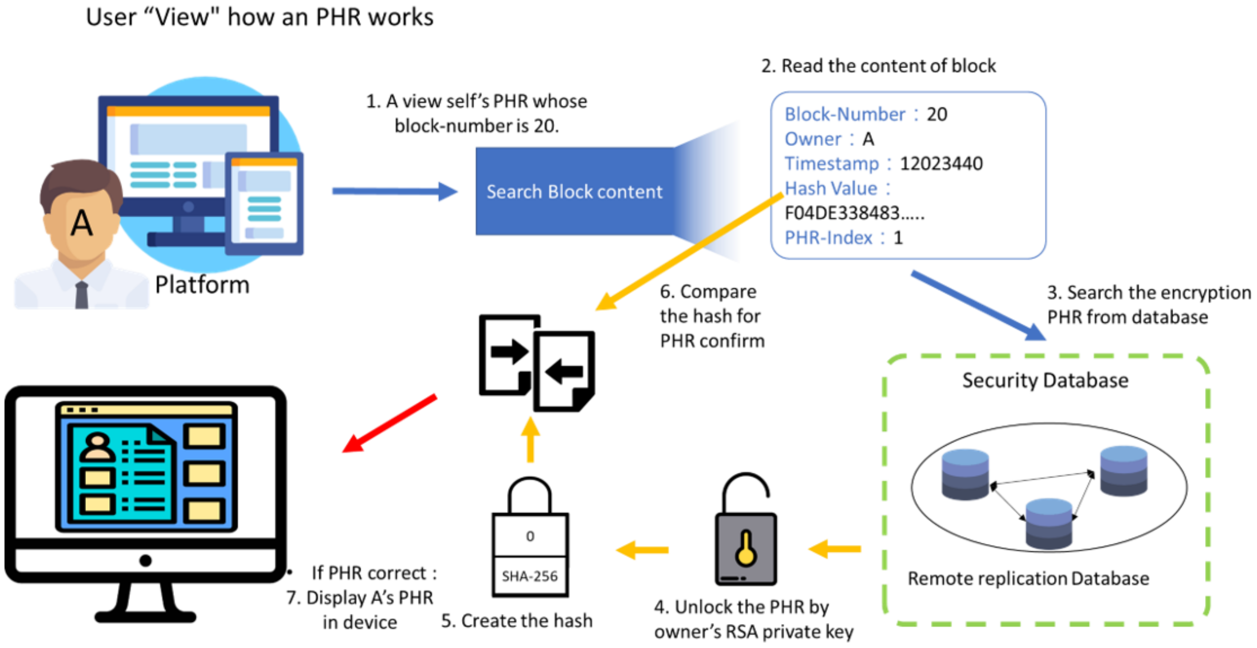


Figure 7. Workflow of a user viewing their own personal health record. PHR: personal health record; RSA: Rivest-Shamir-Adleman.



International Personal Health Record Exchange Implementation Process

The platform can be used at all places where the internet is available. This study used the data format designed in Taiwan MHB as an example to test the system in Asia. MHB contains all the necessary clinical health care data. In Taiwan, 99% of residents can access MHB. Therefore, we chose MHB as an example of PHRs for the AeHIN. The Philippines and Thailand were used as test cases for this study, and 2 of the physician representatives in this study were Dr Alvin in the Philippines and Dr Boonchai in Thailand.

A testing scenario was designed in which a patient from Taiwan travels to Bangkok and the Philippines and suddenly requires medical services. Both the patient and doctors in different countries were registered on this platform. Before the patient would see a doctor in a specific country, authorization to view the PHR would need to be given to the doctor by the patient. For this testing scenario, a patient's PHR with diagnoses of type 2 diabetes mellitus, epilepsy, brain stem stroke, and proteinuria NOS (not otherwise specified) and medication data was designed. The data of testing scenario is descripted in Table 1.

The scenarios consisted of the following scenes:

1. A patient from Taiwan travels to the Philippines.
2. The patient develops a headache and dizziness.
3. The patient goes to see a doctor who has been registered in our platform.
4. Authorization to view the PHR is given to the doctor.
5. The doctor retrieves the patient's PHR from the platform.
6. By viewing the previous PHRs of the patient, the doctor obtains the health profile of the patient and then completes a new diagnosis, treatment, or medication order according to the current status of the patient.
7. A new block is created and the new PHR is stored in the PHR database, if the doctor is willing to upload the new record.

Table 1. The data of testing scenario for international personal health record exchange.

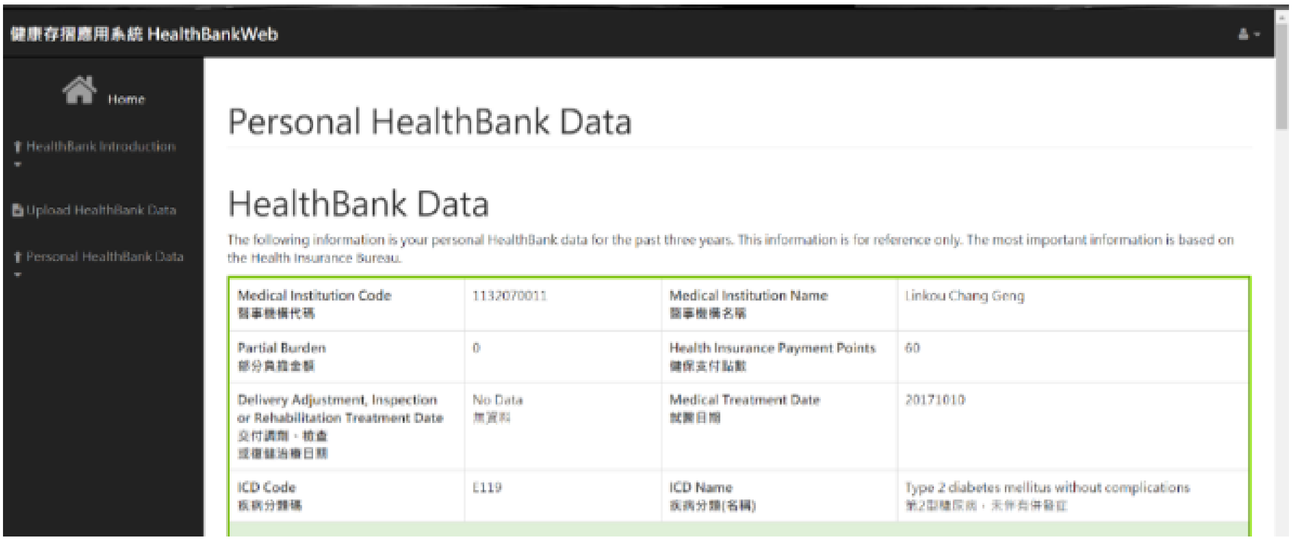
Num	Date	Diagnosis	Medical
1	October 10, 2017	Type 2 diabetes mellitus	Iunaidon Tablets *Yu Sheng*
2	July 16, 2017	Epilepsy	Neurtrol F.C. Tablets 300 mg.
3	May 25, 2017	Brain stem stroke	Cofarin Tab 1 mg
4	May 20, 2017	Brain stem stroke	Cofarin Tab 1 mg
5	January 20, 2017	Proteinuria not otherwise specified	Kaluril Tablets 5 mg

Results

Study Design

This study designed a blockchain-based PHR exchange architecture and management platform for the secure management transfer and sharing of PHR data between patients and medical health care providers. In the PHR management component, the user interface was established; its functions include viewing PHRs for personal health management, sharing PHRs with a doctor, and checking the blockchain content for security. The PHR viewer user interface is shown in Figure 8. MHB was used as an example in this study.

Figure 8. The user interface of a personal health record viewer.



Personal HealthBank Data

HealthBank Data

The following information is your personal HealthBank data for the past three years. This information is for reference only. The most important information is based on the Health Insurance Bureau.

Medical Institution Code	1132070011	Medical Institution Name	Linkou Chang Geng
Partial Burden	0	Health Insurance Payment Points	60
Delivery Adjustment, Inspection or Rehabilitation Treatment Date	No Data	Medical Treatment Date	20171010
ICD Code	E119	ICD Name	Type 2 diabetes mellitus without complications

The Personal Health Record Viewer User Interface of Platform

In Figure 8, the uploaded PHR is displayed. Records are sorted in a time sequence from the latest to the oldest. The display shows the record of each visit, and the patient's medication history. A doctor can view the latest related health record and recent medication status to give the patient the most appropriate diagnosis, while avoiding the problem of adverse reactions between repeated medications or adverse medications.

Blockchain Information in the Platform

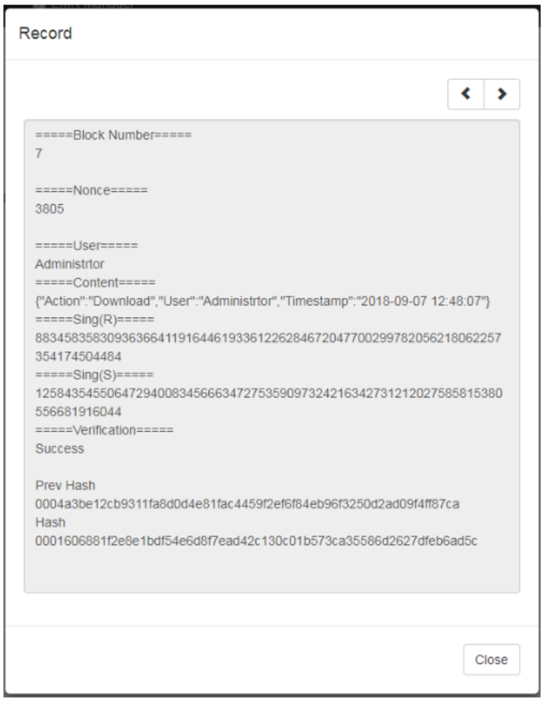
The block content is shown in Figure 9 and includes the time at which the PHR was uploaded, the PHR owner, the PHR hash value, a timestamp, a block hash value, and a pre-hash value. Each block records the previous block location and concatenates to the previous block.

When users upload their personal MHB file, the system automatically converts the file to the FHIR format and transfers it to the security database. The data are then encrypted and uploaded to the blockchain. The users can view the uploaded data records and the contents of the block by uploading the module and obtain a health record for downloading in the FHIR format. A hospital can then upload that data to their system, as long as the system supports the FHIR format.

The blockchain architecture allows users to set their own PHR read permission using the PHR management platform to control who can view their records. The blockchain is used to confirm that the PHR content is correct. The authority control user interface is shown in Figure 10. The simple user interface design ensures that the platform and function are easy to navigate and operate. The design uses 2 columns to display a list of permissions, one of which is a list of trusted participants, and the other is a list of participants to whom the user wishes to grant permission to view their current PHR. When the user wants to grant a doctor permission to view their PHR, they select the doctor from the left-hand column and update the identity.

The blockchain architecture in this study is built by Ethereum, and the blocks are connected by the hash value of each block. The connection diagram is shown in Figure 11.

Figure 9. Block content.



```
=====Block Number=====
7

=====Nonce=====
3805

=====User=====
Administrtor
=====Content=====
{"Action":"Download","User":"Administrtor","Timestamp":"2018-09-07 12:48:07"}
=====Sing(R)=====
88345835830936366411916446193361226284672047700299782056218062257
354174504484
=====Sing(S)=====
12584354550647294008345666347275359097324216342731212027585815380
556681916044
=====Verification=====
Success

Prev Hash
0004a3be12cb9311fa8d0d4e81fac4459f2ef6f84eb96f3250d2ad09f4ff87ca
Hash
0001606881f2e8e1bdf54e6d8f7ead42c130c01b573ca35586d2627dfeb6ad5c
```

Figure 10. Authority control user interface.

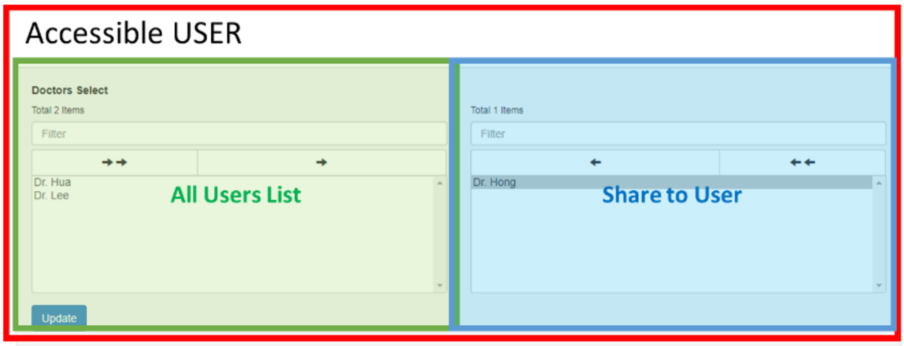
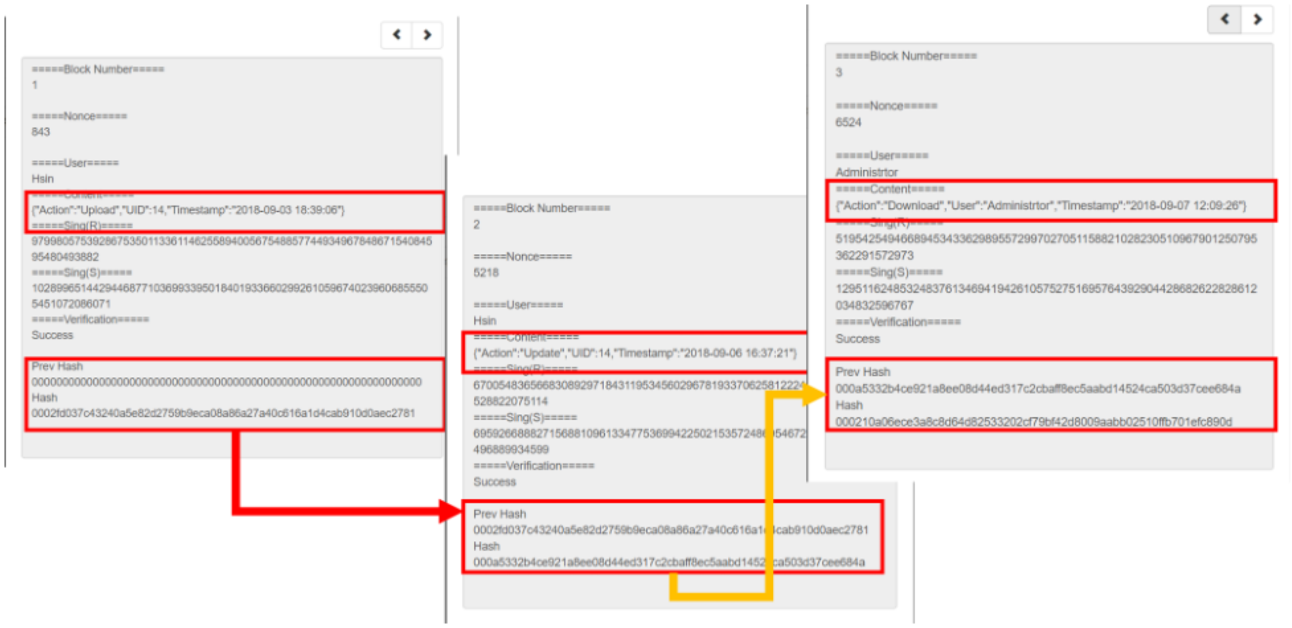


Figure 11. Blockchain connection diagram.



```
=====Block Number=====
1

=====Nonce=====
843

=====User=====
Hsin
=====Content=====
{"Action":"Upload","UID":14,"Timestamp":"2018-09-03 18:39:06"}
=====Sing(R)=====
979980575392867535011336114625589400567548857744934967848671540845
95480493882
=====Sing(S)=====
102899651442944687710369933950184019336602992610596740239606855550
5451072086071
=====Verification=====
Success

Prev Hash
0000000000000000000000000000000000000000000000000000000000000000
Hash
0002fd037c43240a5e82d2759b9eca08a86a27a40c616a1d4cab910d0aec2781
```

```
=====Block Number=====
2

=====Nonce=====
5218

=====User=====
Hsin
=====Content=====
{"Action":"Update","UID":14,"Timestamp":"2018-09-06 16:37:21"}
=====Sing(R)=====
670054836566830892971843119534560296781933706258122224
528822075114
=====Sing(S)=====
6959266888271568810961334775369942250215357248[illegible]54672
496889934599
=====Verification=====
Success

Prev Hash
0002fd037c43240a5e82d2759b9eca08a86a27a40c616a1d4cab910d0aec2781
Hash
000a5332b4ce921a8ee08d44ed317c2cbaff8ec5aabd14524ca503d37cee684a
```

```
=====Block Number=====
3

=====Nonce=====
6524

=====User=====
Administrtor
=====Content=====
{"Action":"Download","User":"Administrtor","Timestamp":"2018-09-07 12:09:26"}
=====Sing(R)=====
519542549466894534336298955729970270511588210282305109679012507950
362291572973
=====Sing(S)=====
129511624853248376134694194261057527516957643929044286826228286120
034832596767
=====Verification=====
Success

Prev Hash
000a5332b4ce921a8ee08d44ed317c2cbaff8ec5aabd14524ca503d37cee684a
Hash
000210a06ece3a8c8d64d82533202cf79bf42d8009aabb02510ffb701efc890d
```

Testing Feedback from Physicians

This study used the MHB data as an example to demonstrate the functioning of the PHR exchange platform and the PHR exchange mechanism based on the blockchain architecture and encryption mechanism, which can ensure PHR storage security and its tamper-proof nature.

The system can more effectively manage self-health records and provide physicians with PHRs as a decision-making reference. The results of this study have been cross-nationally tested in Southeast Asian countries, exchanging PHRs via the AeHIN, and invited physicians from Southeast Asian countries as international participant doctors to allow users to exchange PHRs internationally for appropriate treatment.

The proposed platform was designed to easily share and exchange PHR information electronically. The contents of the PHRs were protected and kept unchanged by the technology of the blockchain architecture. The international standard format of Health Level 7 FHIR was designed in this platform to ensure the interoperability. Doctors could use the platform to upload and download PHR data from different places at any time, thereby allowing PHRs to be exchanged efficiently. Therefore, this platform could increase the accessibility, interoperability, timeliness, and usability of PHRs.

The platform is currently in its testing stage, and there is a low number of users on the network. The users' comments could be summarized as follows:

1. PHRs that are in a standardized format on this platform are a benefit for clinical service.
2. By using the platform, the exchange of PHRs is easy and efficient.
3. The protection offered by the blockchain technology can convince users that the system is secure.
4. Even if the role of the user is that of the platform manager, PHRs still cannot be read without the authorization given by the patient to view the PHR.
5. Personal health management functions can be designed in future work.

Discussion

Potential

Blockchain technology has great potential for electronic health records [13]. The core of the blockchain model ensures that any information involved has nonrepudiation to maintain the correctness of the historical process records [14]. Gary et al [15] Reviewed the current PHR definitions and multiple blockchain architectures for PHR management and found that blockchain technology is a key requirement for the management of consent to use private health data.

Many studies have proposed health applications based on the blockchain technology that can be used in the medical domain to achieve medical record sharing. In 2016, Ekblaw et al [16] created a decentralized medical record management platform that was built on the private network of Ethereum. The platform can only be accessed by authorized users, and blockchain was used to manage authentication, data sharing, and other security functions in the medical field. In the study, when any information was updated on the hospital side, it was uploaded to the blockchain; the platform was synchronized with the patient's database, and the patient would be reminded to update the block. However, patients were unable to upload data themselves, as the data were all still stored in the centralized hospital database. Omar et al [17] used Ethereum's smart contracts and a decentralized application to build a cloud-based PHR system. This system was used to store the PHR of each user and also to ensure the security and integrity of the uploaded data. Private accessible units (PAU) were responsible for all encryption, decryption, uploading of data, searching for data, and verification of data in which users can encrypt data with an encryption key and upload data to the blockchain through a smart contract, which then returns a block-id to the user uploading the data. The user would be responsible for remembering the block-id. To view the data, the user would provide the PAU with the block-id, and the system would automatically return the corresponding block content and decrypt it with the decryption key. This system, however, did not offer the capability of sharing personal medical records or system interoperability.

Peterson et al [18] presented a blockchain-based approach to sharing patient medical data that relies on a single centralized source of trust rather than network consensus to translate data and provides consensus on proof of structural and semantic interoperability. Zhang et al [9] presented a blockchain-based framework FHIR Chain that was designed to fit the technical requirements defined by the Office of the National Coordinator for Health Information Technology interoperability roadmap.

Precision medicine requires the accurate collection and management of all kinds of clinical data. To this end, this study constructed an innovative data storage mechanism, used blockchain technology to ensure the correctness and safety of the PHR data, and combined a security database storage structure with a data verification mechanism to complete data management. A Korean team implemented the blockchain PHR management platform; however, the data transaction time in their study was too long. To allow for the management of queries by a large number of patients, transaction and propagation times must improve [19]. Ahmed et al [20] proposed a blockchain-based emergency access control management system that can protect PHRs using a smart-contract design; however, the system manager can still retrieve real patient data, making privacy issues a concern. The platform designed in this study could offer patient-centered clinical record exchange and decision-making support and allow patients to view and share their own PHRs, as well as manage their health status and apply for medical data using other functions effectively. The platform and architecture could enable the meaningful use of PHRs and promote self-health management. The feasibility was demonstrated by an application test with international users in this study.

An important element of precision medicine is the exchange and management of PHRs and the subsequent provision of personalized medical treatment based on that data during the clinical diagnosis and treatment. This study therefore combined blockchain architecture and data verification methods to effectively solve the problems of data security, storage, and transmission and proposed a hybrid blockchain and data security approach that could enable effective international PHR exchanges. Using the AeHIN's cross-national network environment, PHRs were successfully exchanged, and an international network of medical and health care providers was established to improve the quality of health care and precision medicine internationally.

Principal Findings

The principal findings are as follows:

- A cross-country platform for PHRs was developed in this study. By using this platform, PHRs could be exchanged and shared between different organizations and individuals (doctors, patients, etc) in an efficient manner.
- A PHR platform was built using a blockchain architecture to ensure the security and privacy of health data. Few PHR systems based on blockchain technology have been developed for cross-country data exchange purposes.
- The platform has been tested by several users in different countries in the AeHIN and has shown that it is a suitable platform for PHR sharing and exchange.
- In our design, health data that can be used for precision medicine and can be stored and modeled in the architecture.

Limitations

Currently, our PHR platform is at the prototype stage. Users from limited groups are participating in testing of the platform. However, the hardware architecture will need to be expanded to ensure the good performance of the platform when a large number of users wish to access the system. Furthermore, as the contents of the PHRs will be exchanged and shared by different countries and regions, an international data standard, such as HL7 FHIR, will be required to ensure smooth implementation.

Future Directions

Important points regarding the comparison with prior work are as follows:

- Precision medicine is the future trend of health care and must be based on PHRs. Our PHR platform not only enables PHRs to be shared between countries but also creates space for future functions of precision medicine.
- Blockchain technology ensures data security and privacy and has been successfully used in financial data management systems.
- A cross-country medical care architecture must be developed in the present busy international activities.

Conclusions

On the basis of the blockchain technology, it is possible to remove all limitations to patients' ability to copy and transfer their own health records to other health service providers [21]. After data are uploaded in the blockchain, the block can guarantee that the records cannot be modified by anyone [22]. The PHRs are stored in a decentralized network; therefore, it is impossible to steal PHR data or hack the system illegally [21]. In addition to improved health record sharing and analysis, data sharing will be secured and privacy will be protected [23].

In addition, the blockchain technology is essential for future precision medicine applications. Through the blockchain architecture, the data required by precision medicine can be integrated from different sources. In addition to using blockchains as a ledger for patient care data, they can also be used to store various types of health care–related data, such as precision medical data and genomic data [24], health care plan data, patient-centered data [25], clinical trial data [26], medication supply chain data, and biomarker data [27-29]. In this study, we implemented a cross-country platform for PHRs. By using this platform, PHRs can be exchanged and shared between different organizations in an efficient manner. The platform has been tested by several users in different countries in the AeHIN and has been shown to be a suitable platform for PHR sharing and exchange. With our design, the health data that can be used for precision medicine can be stored and further modeled in the architecture. The security and privacy of PHRs can also be ensured by the features of blockchain technology, such as distributed node consensus algorithms, data transmission cryptography, and a decentralized network of smart contracts. However, an international standard, such as FHIR, will be required to ensure the PHR contents are internationally compatible.

Acknowledgments

This project has received funding from the Ministry of Science and Technology, Taiwan, under the project no. 108-3011-F-075-001 and Ministry of Education, Taiwan, under the project no. 107EH12-22.

Authors' Contributions

The work presented in this paper was carried out in collaboration among all authors. HL and CH conceptualized the study and study design and also designed the architecture of the system. HL, HK, and JU carried out literature review and system analysis. HK put a lot of effort in the implementation of the system. HL drafted the manuscript, and CH made significant revisions. JU, BK, and AM remotely tested the system. CH, JU, BK, AM, and LC supervised the methods of the implementation on a cross-country platform and suggested valuable improvements. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Farmer C, Fenu E, O'Flynn N, Guthrie B. Clinical assessment and management of multimorbidity: summary of NICE guidance. Br Med J 2016 Sep 21;354:i4843. [doi: 10.1136/bmj.i4843] [Medline: 27655884]
2. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
3. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]
4. AHIMA e-HIM Personal Health Record Work Group. American Health Information Management Association. 2005. Role of the Personal Health Record in the EHR (2010 update) - Retired URL: https://library.ahima.org/doc?oid=103209#.XnGsmagzaM8 [accessed 2020-03-18]
5. American Health Information Management Association. New York: Markle Foundation; 2003. Connecting for Health: A Public-Private Collaborative(Report) URL: http://bok.ahima.org/PdfView?oid=76138 [accessed 2020-03-18]
6. Nakamoto S. Bitcoin. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System URL: https://bitcoin.org/bitcoin.pdf [accessed 2020-03-18]
7. Halamka JD, Lippman A, Ekblaw A. Harvard Business Review. 2017 Mar 3. The Potential for Blockchain to Transform Electronic Health Records URL: https://hbr.org/2017/03/the-potential-for-blockchain-to-transform-electronic-health-records [accessed 2017-03-03]
8. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
9. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: applying blockchain to securely and scalably share clinical data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
10. NIST Computer Security Resource Center. Gaithersburg: NIST FIPS; 2002 Aug 1. Secure Hash Standard URL: https://csrc.nist.gov/csrc/media/publications/fips/180/2/archive/2002-08-01/documents/fips180-2withchangenotice.pdf [accessed 2020-03-18]
11. Dang Q. National Institute of Standards and Technology. Gaithersburg: US Department of Commerce, National Institute of Standards and Technology; 2012 Aug. Recommendation for Applications Using Approved Hash Algorithms URL: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-107r1.pdf [accessed 2020-03-18]
12. Rivest RL, Shamir A, Adleman L. A method for obtaining digital signatures and public-key cryptosystems. Commun ACM 1978;21(2):120-126. [doi: 10.1145/359340.359342]
13. Dimitrov DV. Blockchain applications for healthcare data management. Healthc Inform Res 2019 Jan;25(1):51-56 [FREE Full text] [doi: 10.4258/hir.2019.25.1.51] [Medline: 30788182]
14. Mendes D, Rodrigues I, Fonseca C, Lopes M, García-Alonso JM, Berrocal J. Anonymized distributed PHR using blockchain for openness and non-repudiation guarantee. In: Méndez E, Crestani F, Ribeiro C, David G, Lopes J, editors. Digital Libraries for Open Knowledge. Cham: Springer; 2018:381-385.

15. Leeming G, Cunningham J, Ainsworth J. A ledger of me: personalizing healthcare using blockchain technology. Front Med (Lausanne) 2019;6:171 [FREE Full text] [doi: 10.3389/fmed.2019.00171] [Medline: 31396516]
16. Ekblaw A, Azaria A, Halamka JD, Lippman A. A Case Study for Blockchain in Healthcare: 'MedRec' Prototype for Electronic Health Records and Medical Research Data. In: Proceedings of the 2016 IEEE Open & Big Data Conference. 2016 Presented at: IEEE BigData'16; August 22-24, 2016; Washington DC URL: https://www.healthit.gov/sites/default/files/5-56-onc_blockchainchallenge_mitwhitepaper.pdf
17. Omar AA, Bhuiyan MZ, Basu A, Kiyomoto S, Rahman MS. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Gener Comput Syst 2019;95:511-521. [doi: 10.1016/j.future.2018.12.044]
18. Peterson K, Deeduvanu R, Kanjamala P, Boles K. A Blockchain-Based Approach to Health Information Exchange Networks. In: Proceedings of the 2016 NIST Blockchain & Healthcare Workshop. 2016 Presented at: NIST'16; September 26-27, 2016; Gaithersburg, MD URL: https://www.healthit.gov/sites/default/files/12-55-blockchain-based-approach-final.pdf
19. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. J Med Internet Res 2019 Feb 8;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
20. Rajput AR, Li Q, Ahvanooey MT, Masood I. EACMS: emergency access control management system for personal health record based on blockchain. IEEE Access 2019;7:84304-84317. [doi: 10.1109/access.2019.2917976]
21. Ivan D. HealthIT. 2016 Aug. Moving Toward a Blockchain-Based Method for the Secure Storage of Patient Records URL: https://www.healthit.gov/sites/default/files/9-16-drew_ivan_20160804_blockchain_for_healthcare_final.pdf [accessed 2020-03-18]
22. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
23. Linn LA, Koo MB. HealthIT. Blockchain for Health Data and Its Potential Use in Health It and Health Care Related Research URL: https://www.healthit.gov/sites/default/files/11-74-ablockchainforhealthcare.pdf [accessed 2020-03-18]
24. McKernan KJ. The chloroplast genome hidden in plain sight, open access publishing and anti-fragile distributed data sources. Mitochondrial DNA A DNA Mapp Seq Anal 2016 Nov;27(6):4518-4519. [doi: 10.3109/19401736.2015.1101541] [Medline: 26486305]
25. Goldwater JC. HealthIT. Bethesda: National Quality Forum; 2016 Aug 8. The Use of a Blockchain to Foster the Development of Patient-Reported Outcome Measures URL: https://www.healthit.gov/sites/default/files/6-42-use_of_blockchain_to_develop_proms.pdf [accessed 2020-03-18]
26. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]
27. Taylor P. Securing Industry. 2016 Apr 27. Applying Blockchain Technology to Medicine Traceability URL: https://www.securingindustry.com/pharmaceuticals/applying-blockchain-technology-to-medicine-traceability/s40/a2766/#.XmoIhpMzZ0t [accessed 2019-09-18]
28. Jenkins J, Kopf J, Tran BQ, Frenchi C, Szu H. Bio-Mining for Biomarkers With a Multi-Resolution Block Chain. In: Proceedings of the 2015 SPIE Sensing Technology + Applications conference. 2015 Presented at: SPIE DSS'15; April 20-24, 2015; Baltimore. [doi: 10.1117/12.2180648]
29. IBM Global Business Services Public Sector Team. HealthIT. Bethesda: IBM Global Business Services Public Sector Team; 2016 Aug 8. Blockchain: The Chain of Trust and Its Potential to Transform Healthcare - Our Point of View URL: https://www.healthit.gov/sites/default/files/8-31-blockchain-ibm_ideation-challenge_aug8.pdf [accessed 2020-03-18]

Abbreviations

AeHIN: Asia eHealth Information Network
ECC: elliptic curve cryptography
FHIR: Fast Healthcare Interoperability Resource
EMR: electronic medical record
MHB: My Health Bank
NHI: National Health Insurance
PAU: private accessible units
PHR: personal health record
PoA: proof of authority
RSA: Rivest-Shamir-Adleman

Edited by G Eysenbach; submitted 21.10.19; peer-reviewed by JT te Gussinklo, B Vaes; comments to author 23.12.19; revised version received 14.02.20; accepted 22.02.20; published 09.06.20.

Please cite as:
Lee HA, Kung HH, Udayasankaran JG, Kijsanayotin B, B Marcelo A, Chao LR, Hsu CY
An Architecture and Management Platform for Blockchain-Based Personal Health Record Exchange: Development and Usability Study
J Med Internet Res 2020;22(6):e16748
URL: https://www.jmir.org/2020/6/e16748
doi:10.2196/16748
PMID:32515743

Original Paper

Inferring Sexually Transmitted Infection Risk From Attractiveness in Online Dating Among Adolescents and Young Adults: Exploratory Study

Tamar Krishnamurti[1], PhD; Alexander L Davis[2], PhD; Baruch Fischhoff[2,3], PhD

[1]Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[2]Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, United States
[3]Institute for Politics and Strategy, Carnegie Mellon University, Pittsburgh, PA, United States

Corresponding Author:
Tamar Krishnamurti, PhD
Division of General Internal Medicine
University of Pittsburgh
200 Meyran Avenue
Parkvale Building Suite 200
Pittsburgh, PA, 15213
United States
Phone: 1 412 692 4855
Email: tamark@pitt.edu

Abstract

Background: Sexually transmitted infection (STI) rates are on the rise among adolescents and young adults in the United States. With the popularity of online dating, adolescents and young adults must increasingly rely on limited cues to make initial judgments about potential sexual partners, including judgments about STI risk.

Objective: This study aimed to assess whether in the context of online dating, an attractiveness heuristic would be used for STI risk assessment. We hypothesized that consistent with research on halo effects, decision makers would judge more attractive people to be less likely to have STIs.

Methods: In a survey experiment, we asked participants to determine which individual in each of 20 sets of paired photographs was enrolled in a personals website for people with publicly disclosed STIs.

Results: Despite financial incentives for accuracy and high levels of self-confidence in their judgments, participants performed no better than chance at identifying individuals with self-reported STIs. Contrary to our hypothesis, however, more attractive people were judged as being more likely to have an STI. This relationship appears to be mediated by inferences regarding the target individual's sexual behavior, with more attractive individuals considered to have more partners.

Conclusions: On showing adolescents and young adults photographs offering no diagnostic information about STIs, they appeared to use attractiveness as a cue for sexual risk, which was mediated by the belief that attractive individuals have more sexual opportunities. Health care providers may wish to address this heuristic process among their adolescent patients in discussions about sexual health.

(J Med Internet Res 2020;22(6):e14242) doi:10.2196/14242

KEYWORDS

risk perception; heuristics; sexually transmitted infections; online dating; dating apps; adolescents; sexual health; attractiveness; halo effect

Introduction

In the United States, sexually transmitted infections (STIs) are increasingly among the most commonly reported diseases, with the total cases of STIs reaching a historical high in 2017 [1,2]. Young adults and adolescents are at particular risk [2]. Public health officials have suggested that online dating and use of dating apps may play critical roles in this burgeoning problem. Over half of the users of popular dating apps are in the highest STI age bracket (under 25 years) [3]. Some young adults have a belief that they can "just tell" by looking whether a potential partner has an STI [4,5]. In a recent study, among a sample of young adults using dating apps, those who transitioned from a profile view to having an in-person date had higher self-reported rates of risky sexual behavior than those who did not transition to face-to-face interactions [6]. Here, we examine one aspect of the process that can lead to sexual contact and STI risk, which is the inference that young people make regarding potential partners' STI risk according to their personal appearance.

The interfaces of popular dating sites, such as Tinder, feature photographs of potential partners and encourage scrolling quickly through them and swiping "right" on appealing individuals for further examination [7]. Beneath the photographs are profiles that might be consulted for those who pass such initial screening. As it is socially awkward to ask about STI status [8] and dating apps rarely provide this information [9], young people must rely on intuition when making judgments about potential partners' STI risk. Attractiveness can be a valid cue for predicting disfiguring STIs (eg, advanced syphilis). On the other hand, for the far more frequent cases of asymptomatic STIs, attractiveness provides no directly relevant information. However, attractiveness could provide indirectly relevant information if it is correlated with risk factors, such as number of sexual partners, frequency of sex, access to health care, and use of condoms.

Existing research offers conflicting evidence regarding the roles of attractiveness judgment in inferences regarding STI risk. Some studies found that when asked explicitly, young people expect more attractive individuals to have greater STI risk, reasoning that they will have more opportunities for varied sexual activities and be more promiscuous [10,11]. A large body of studies, however, suggested the opposite. For example, one study found that male respondents provided lower estimates of STI risk for women described as "attractive" in thumbnail personality sketches [12]. Another found an inverse relationship between how "attractive" various qualities were perceived in a potential partner and how "risky" those qualities were judged to be [13]. These studies may involve a "halo effect," whereby one positive feature, such as physical attractiveness, encourages other positive perceptions [14,15]. Indeed, some evidence suggests that motivated rather than deductive reasoning may color subjects' judgments about sexual purity, when self-justifying their failures to use condoms with more attractive partners [16,17]. There is also strong evidence that people seen as more attractive are also viewed as more intelligent, academically and socially competent, politically knowledgeable, and cooperative [18-23]. This halo effect is present for visual judgments of both male and female individuals and remains in place once individuals have interacted with one another. If sexual risk is within the halo of attractiveness, young people may infer a lower STI risk from potential partners' physical attractiveness in online dating contexts.

In this study, we created such a context experimentally in order to examine whether attractiveness is a cue for risk when young people make judgments from photographic cues alone. We also examined their confidence and accuracy. We posited that people making quick judgments about potential dating partners in online dating profiles would apply an *attractiveness heuristic*. Namely, they would perceive attractive individuals as less likely to have an STI.

Previous studies of perceived STI risk have typically asked participants to make inferences from multiple (sometimes contradictory) cues [10,12] or have asked them to evaluate attractiveness and risk simultaneously, with explanation of their inferences [10-13,24]. Our study adds to this research by examining rapid judgments based on visual cues alone, with no prompt for reasoned inferences. Its results have implications for online dating contexts and, more generally, for the connections between fast and slow thinking [25].

We asked participants to judge the likelihood that individuals in photographs have STIs under conditions that should automatically evoke judgments of the targets' attractiveness (making snap judgments based on rapidly displayed pairs of photographs). The judgment is which of the two pictures has been drawn from a website for people with self-disclosed STIs. To develop this test set of photographs, we had a separate sample rate the target in each photograph in terms of attractiveness and several factors that might mediate the relationship with STI risk (number of sexual partners [as a proxy for STI exposure], intelligence, frequency of condom use, and decision-making competence [potentially protecting from exposure]).

Methods

Baseline Photograph Data Set

To characterize the photographs used in the experiment, we had a separate pretest sample of 125 heterosexual individuals (aged 18-25 years; 39% female), who were recruited from a university student pool, rate opposite-sex photographs for attractiveness and several risk-related characteristics. Raters were recruited for a web-based survey in which they would be "asked to make judgments about individuals," and they received university course credit for participating. They were eligible for the rating study if they self-identified as heterosexual, were at least 18 years of age, and self-reported not being in a romantic relationship. Photographs were drawn from profiles of residents across the United States. All photographs showed someone who identified as being between the ages of 18 and 25 years and heterosexual in their original online dating profile. Each individual was photographed looking directly at the camera. Among photographs of both male and female individuals, there were approximately 80% Caucasian people. Photographs were cropped square (1:1 aspect ratio) to show only the neck and face and to minimize surrounding visuals. Photographs were all in color and were selected by two independent research assistants as having a pleasant or neutral facial expression.

All photographs were publicly available, and use of the photographs complied with the terms of service of the websites at the time that the stimulus photographs were gathered. Although the photographs, by their nature, identified the individuals depicted, they were obtained from national sources; hence, there was a very low probability of including individuals known to the participants. The sources of the photographs were not revealed to the participants.

There were 96 photographs in total (48 photographs of male individuals and 48 of female individuals). Each photograph was rated by 10 raters. Each rater judged 12 unique, randomly selected, opposite-sex photographs. Photographs were rated on physical attractiveness (1 [very unattractive] to 7 [very attractive]), according to the approach in other studies on facial attractiveness [26,27]. Photographs were also rated on the following three risk-related characteristics drawn from prior studies [15], which could be protective against sexual risk: (1) *intelligence* (1 [not at all intelligent] to 7 [very intelligent]), (2) *competence* (1 [foolish] to 7 [sensible]), and (3) *condom use with a new partner* (1 [never] to 7 [always]). Additionally, photographs were rated on the following factor that could increase sexual risk: *likelihood of multiple sexual partners* (1 [not at all likely] to 7 [very likely]). Raters used the entire 7-point Likert scale. The average attractiveness score in ratings of photographs of male individuals ranged from 1.75 to 5.0 and in ratings of photographs of female individuals ranged from 1.35 to 5.50. Controlling for the website from which the photographs were drawn, photographs of male individuals were rated as slightly less attractive than those of female individuals (t_{77}=2.11, *P*=.04). There were no significant differences in the attractiveness ratings of photographs from each dating website (*P*=.11; the average attractiveness ratings for STI website photographs ranged from 1.61 to 5.00 and for non-STI website photographs ranged from 1.35 to 5.50).

Descriptive statistics for the pretest sample's ratings of the photographs that were ultimately retained to be used in the research study (10 pairs of male individuals and 10 pairs of female individuals; rated on a scale from 1 to 7) along with the correlations between them can be found in Multimedia Appendix 1. Individuals judged as more attractive were given higher ratings on the three protective factors (intelligence, frequency of condom use, and competent decision-making) and one risk factor (multiple sexual partners).

Study Participants

Our study included 87 participants (55 male and 32 female participants) recruited from a private university student participant pool using online postings and recruited on the street in a high foot-traffic neighborhood housing multiple universities (both public and private) with a sign posted outside a research laboratory. The posting stated that participants were being recruited for "a study to understand individual decision-making" and could participate if they met the inclusion criteria of being at least 18 years of age and not participating in rating the baseline photograph data set. Participants recruited from the university student participant pool were emailed a link to the study. Participants recruited on the street outside the laboratory completed the study on a computer inside a private cubicle.

Study Procedure

Participants were shown 20 pairs of photographs (10 pairs of male individuals and 10 pairs of female individuals) drawn from the prerated photograph set. One photograph in each pair was drawn from a personals website for people who have publicly disclosed an STI. The other photograph was drawn from a dating website without that disclosure. Participants were told about the two websites and the photograph sampling procedure. They were also asked to assume that people from the non-STI disclosure website had the same rate of STIs as the general population. For each photograph pair, participants were asked which photograph was from the STI website and about their confidence in the choice (from 50% [chance] to 100% [certainty]). For each pair, one photograph was randomly sampled from each site and assigned randomly to the right- or left-hand side.

Participants received US $5 in compensation for their time or university course credit. All participants received an additional US $0.25 for each correct response. This amount was selected to provide an incentive for accuracy without compromising the rapid judgment process. Such incentives have been found to increase attention without reducing errors attributable to heuristic use [28].

For those who completed the study online, the primary researcher evaluated responses for accuracy and emailed the participants about their payment, which was collected from another researcher. Although this researcher could infer the number of correct responses from the payment amount, there was no information about which stimuli a participant had seen. The data were fully deidentified upon payment and prior to the analysis.

After completing the task, participants answered questions about their relationship and sexual history, including binary response questions, such as *Are you currently in a romantic relationship?*, *Are you currently sexually active?*, and *Have you ever had a "one-night stand"?*, and numeric response questions, such as *How many sexual partners have you had in total?* and *How many times have you had sex in your lifetime?*, as well as a categorical sexual orientation question.

The study, including the acquisition and use of stimuli, was approved by the Institutional Review Board (IRB) of Carnegie Mellon University, which designated the study as posing minimal risk. The IRB did not require informed consent from the individuals in the stimuli as, at the time, all photographs were publicly available, with no requirement to create an account to view or download them for research purposes. To ensure the privacy of the individuals in the photographs, we have not made their images publicly available. Our code, survey, and stimuli rating data set are available publicly [29].

Results

Participant Characteristics

Of the 87 participants, 74 identified as heterosexual, 12 as homosexual or bisexual, and one did not respond. The participants ranged in age from 18 to 56 years, with a mean age of 22.8 (SD 7.8) years. Table 1 presents the participants' self-reported demographic characteristics.

Among the 87 participants, 79 (91%) responded to the question about prior sexual activity, with 50% (26/52) of male participants and 22% (6/27) of female participants reporting none.

Table 1. Participants' characteristics.

Variable	Value (N=87), n (%)
Gender	
Male	55 (67%)
Female	32 (33%)
Age[a]	
18-24	65 (75%)
25-34	10 (11%)
≥35	5 (6%)
Currently in a romantic relationship	
Yes	35 (40%)
No	45 (52%)
Currently sexually active	
Yes	42 (48%)
No	38 (44%)
Total lifetime instances of sex (eg, penile-vaginal, oral, and anal)	
0 instances	36 (41%)
1-10 instances	13 (15%)
>10 instances	31 (36%)
History of "one-night stands"	
Never	67 (77%)
At least once	13 (15%)

[a] Of the 87 participants, 7 (8%) did not provide complete demographic data.

Descriptive Statistics

Sample Size Considerations

Each of the 87 participants made 20 judgments about which individual in a pair of photographs was more likely to have an STI. Our effective sample size is somewhere between 87 observations (if each participant's responses are perfectly correlated) and 1740 observations (20 × 87). Assuming an effective sample size of 87, the statistical power is 0.15 to detect a small effect (r=0.1), 0.81 to detect a medium effect (r=0.3), and 0.99 to detect a large effect (r=0.5).

STI Identification

The mean percentage of correct STI identifications was 47% (9.85/20). Participants' mean confidence in their judgments was 67.2% (SD 16.0%). Thus, in aggregate, participants were overconfident, expecting more correct identifications than were observed.

Attractiveness Heuristic

To test for the use of an attractiveness heuristic in our primary experiment, we estimated a series of binary mixed logit models predicting the probability of participants predicting that an image was from the STI website as a function of the difference in the mean attractiveness ratings of the two images (for the pretest sample) and controlling for the "ground truth" of a self-disclosed STI. The following three models were created:

Models 1-3 in Table 2 are indexed by participant i and photograph pair j. The dependent variable y_{ij} was coded as 1 if participant i selected the photograph on the right from photograph pair j and as 0 otherwise. Each model used the logit function to relate p_{ij}, the modeled probability that participant i selects the right photograph in pair j as having an STI, to characteristics of the photograph and the participant. In model 1 (M_1), we included an intercept β_{0i} to allow for individual-specific tendencies to select the photograph on the right and β_{1i} times the difference in attractiveness between the right and left photographs for pair j to capture individual-specific tendencies to use attractiveness as a cue for STI risk. The coefficients β_{0i} and β_{1i} are assumed to be drawn from a multivariate normal distribution with mean 0 and an unrestricted variance-covariance matrix. Model 2 (M_2) adds an indicator variable for whether the photograph on the right involves an STI for photograph pair j, with a corresponding coefficient β_{2i} that captures individual i's ability to detect the presence of STI from the photographs. Model 3 (M_3) adds additional individual-level covariates. For all models, we used the

Nelder-Mead optimizer from the lme4 package in R for estimation.

As seen in Table 2, model 1 found that as the difference in attractiveness increases, participants are *more* likely to identify the more attractive individual as having been drawn from the STI website. This relationship held when, in model 2, the actual website was added to the equation, with that information not adding significant predictive power (consistent with participants predicting the actual website at chance level). Model 3 showed that the pattern held after adding four participant-level variables, none of which added predictive power (gender, age, gender match with the target individual, and the participant's reported number of lifetime sexual partners).

Across all participants, a greater difference in attractiveness between the two photographs being judged was associated with a higher probability of the participant predicting that the more attractive individual was from the STI website (*Z*=2.08, *P*=.04).

Table 2. Mixed logit models predicting whether the target was drawn from the sexually transmitted infection disclosure website, according to attractiveness ratings, the actual website, and personal characteristics.

Variable	Model 1 (1738 observations, 87 participants)			Model 2 (1738 observations, 87 participants)			Model 3 (1578 observations, 79 participants)		
	Estimate	SE	SD	Estimate	SE	SD	Estimate	SE	SD
Intercept	−0.11	0.07	0.40	−0.08	0.08	0.41	−0.04	0.25	0.19
Attractiveness difference score	0.10[a]	0.05	0.20	0.10[a]	0.05	0.20	0.11[a]	0.05	0.21
Self-disclosed STI[b]	—[c]	—	—	−0.07	0.10	0.02	−0.07	0.10	0.07
Gender	—	—	—	—	—	—	0.12	0.11	—
Age	—	—	—	—	—	—	−0.001	0.01	—
Target matched on gender	—	—	—	—	—	—	0.05	0.10	—
Total number of sexual partners	—	—	—	—	—	—	0.02	0.05	—

[a] *P*<.05.

[b] STI: sexually transmitted infection.

[c] Not entered into the regression.

Role of Risk-Relevant Characteristics in STI Risk Judgments

We used binary mixed logit models to assess the role of other features of the photographs, using pretest sample ratings (Multimedia Appendix 1) added to models 4-6 (Table 3). As seen in Table 3, model 4 found that the difference in target-perceived intelligence added no predictive power. Model 5 found that the difference in the ratings of condom use added predictive power, without affecting the relationship with attractiveness. Model 6 found that the difference in the ratings of the targets having multiple partners added predictive power as well, with attractiveness no longer playing a role.

Table 3. Mixed logit models predicting whether the target was drawn from the sexually transmitted infection disclosure website, according to attractiveness ratings, the actual website, and photograph ratings.

Variable	Model 4 (1738 observations, 87 participants)			Model 5 (1738 observations, 87 participants)			Model 6 (1738 observations, 87 participants)		
	Estimate	SE	SD	Estimate	SE	SD	Estimate	SE	SD
Intercept	−0.04	0.09	0.44	−0.05	0.09	0.40	−0.06	0.08	0.39
Attractiveness difference score	0.13[a]	0.05	0.20	0.15[a]	0.05	0.21	−0.03	0.06	0.28
Self-disclosed STI[b]	−0.15	0.11	0.10	−0.14	0.10	0.07	−0.10	0.10	0.09
Intelligence difference score	−0.11[c]	0.05	0.12	—[d]	—	—	—	—	—
Condom use difference score	—	—	—	−0.35[a]	0.07	0.21	—	—	—
Multiple partners difference score	—	—	—	—	—	—	0.38[a]	0.07	0.21

[a] *P*<.01.

[b] STI: sexually transmitted infection.

[c] *P*<.05.

[d] Not entered into the regression.

Discussion

Principal Findings

Young adults and adolescents engaging in online dating have to generate quick intuitive judgments when making choices about their interactions with potential sexual partners. Like other decision makers, when they lack statistical estimates, they may rely on heuristics to judge risk [30]. Online dating invites such heuristic judgments in decisions about engaging others as potential sexual partners. In this study, we examined the potential role of an attractiveness heuristic in sexual risk judgment by asking participants to predict which of two photographs came from a website for individuals with self-reported STIs. We assumed that participants would automatically evaluate the target's attractiveness and then apply the attractiveness heuristic to that judgment.

We did find that attractiveness predicts judgments about STI risk. However, the direction of the association was contrary to our prediction that STIs would seem less likely in more attractive individuals (ie, observing a "halo effect"). Instead, attractiveness appeared to be used as a cue for higher sexual risk. Analyses incorporating other variables led us to an alternative post hoc explanation, which is consistent with research findings that attractive people are perceived as more sexually promiscuous [10,11]. In our study, the relationship between judgments of attractiveness and STI risk appeared to be mediated by judgments of target individuals' numbers of partners; more attractive individuals are considered to have more partners and hence a greater STI risk. Thus, the archetype of an attractive individual may not be associated with "purity" as much as with "opportunity." It is also possible that the confidence judgments evoked more analytical and less heuristic thinking, producing more measured inferences about sexual "opportunity," even in a context designed to encourage snap judgments.

Whatever processes guided their judgments, participants were unable to predict which photographs were drawn from the website with STI disclosure, despite incentives for accuracy. Moreover, they showed overconfidence that is typical of difficult tasks [31]. Their predictions were related to the normatively valid cue of whether the target was rated (by the pretest sample) as someone likely to use condoms and, perhaps, with greater intelligence (Table 3). However, those perceptions appeared to reflect judgments of attractiveness, suggesting sound inferences based on unsound assumptions (Multimedia Appendix 1). As seen in model 6 (Table 3), inferences reflecting attractiveness appear to be subsumed by inferences regarding multiple sexual partners. Given how we created the stimulus set, attractiveness was not a valid cue for predicting a photograph's origin and yet it played a role in participants' judgments.

Limitations

Our study had several notable limitations. First, photograph pretesting was limited to judgments by male and female individuals identifying as heterosexual. Thus, it is possible that the attractiveness ratings for these photographs did not reflect the perceptions of the mixed heterosexual and bisexual participants in our experimental study sample. However, there is evidence that judgments of attractiveness of same and other-sex individuals differ by gender but not sexual orientation [32]. To the best of our knowledge, there is no analogous evidence regarding judgments of STI risk and our sample size did not allow subanalyses by sexual orientation. Understanding how such judgments relate to sexual orientation is a topic for future research, and it is particularly relevant given the popularity of dating apps among gay, bisexual, and other men who have sex with men, as well as their higher risk of contracting STIs [2].

Our convenience sample had a gender imbalance, perhaps reflecting greater interest among male participants (65%) in an experiment about "online dating." As such, our results are more generalizable to those who are interested in online dating. Within the constraints of our sample, STI predictions were unrelated to gender, gender match with photographs, age, or self-reported STIs (Table 2).

While our sample consisted primarily of college students recruited through a university student pool and other young people recruited in a public setting, we did have a sizable proportion (15/87, 17%) of participants aged ≥25 years. Our results did not differ by age group; however, the lack of age-based exclusion in our recruitment procedures should be noted when interpreting the findings in the pediatric context. Moreover, by advertising the topic, as well as recruiting some participants in person, we may have biased the recruited sample.

According to participants' self-reports, 41% (36/87) had no sexual encounters and 36% (31/87) had more than 10 encounters. The large representation of sexually inexperienced individuals is unlike much sex-related research [33-35] and higher than reports in national samples of this age group [36]. We cannot predict how this or other selection processes might have influenced our results. We attempted to focus participants on the images, rather than the individuals in them, by presenting targets who participants would never meet and by providing financial incentives for accuracy. This approach deprived participants of both valid and invalid cues available in everyday life, the most important of which may be the website on which the photograph appeared. Thus, our results suggest a form of heuristic thinking that young people may use, without indicating its power or prevalence.

Conclusions

Romantic interest in a target has been found to be driven primarily by sexual attractiveness [35]. For many people, the attractiveness of a potential partner's online image affects whether to proceed in engaging with them [37]. Using an experimental task that involved rapid judgments of photographs, we found that judgments of STI risk were related to both relevant risk factors and the irrelevant cue of attractiveness. The role of attractiveness appears to have been mediated by the rapid inference that attractive people have more sexual partners and hence greater STI risk. Although participants predicted STI risk at chance levels (47%), they had moderate confidence in their predictive ability (mean confidence 67%). Reliance on heuristic judgments of risk from visual cues alone, paired with a misplaced confidence in the ability to identify risk among others, could contribute to the higher rates of STIs among those who use online dating sites to initiate sexual encounters [38-40].

These findings may help inform conversations between health care providers and their young patients, providing content for STI counseling that many are eager to provide [41]. Such conversations offer one way to address the medical community's call for more proactive approaches to stem the rise in STI rates [42]. Those conversations may include discussions on the validity of the cues available in online dating apps. This research is a formative step toward understanding young people's inference processes when using apps that play central roles in many of their lives. Further research should examine heuristics that might better inform the snap judgments of sexual risk that accompany "swiping left or right."

Acknowledgments

The authors gratefully acknowledge the support of an National Science Foundation Graduate Research Fellowship for the costs of collecting data and compensating research assistants. TK's time in preparing the manuscript was supported by National Institute of Health (NIH) KL2 TR001856. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. TK and AD are cofounders of Naima Health LLC, which develops digital health tools to engage patients in clinical care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive statistics and Spearman rank correlations for pretest judgments of target photographs.
[DOCX File , 18 KB - jmir_v22i6e14242_app1.docx]

References

1. Pearson WS, Peterman TA, Gift TL. An increase in sexually transmitted infections seen in US emergency departments. Prev Med 2017 Jul;100:143-144. [doi: 10.1016/j.ypmed.2017.04.028] [Medline: 28455221]
2. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (U.S.). Division of STD Prevention. Centers for Disease Control and Prevention. 2019 Oct 01. Sexually transmitted disease surveillance 2018 URL: https://stacks.cdc.gov/view/cdc/79370 [accessed 2020-01-30]
3. Sawyer AN, Smith ER, Benotsch EG. Dating Application Use and Sexual Risk Behavior Among Young Adults. Sex Res Soc Policy 2017 Sep 6;15(2):183-191 [FREE Full text] [doi: 10.1007/s13178-017-0297-6]
4. Wyn J. Young women and sexually transmitted diseases: the issues for public health. Aust J Public Health 1994 Mar;18(1):32-39. [doi: 10.1111/j.1753-6405.1994.tb00191.x] [Medline: 8068790]
5. Clark LF, Miller KS, Harrison JS, Kay K, Moore J. The role of attraction in partner assessments and heterosexual risk for HIV. In: Oskamp S, Thompson SC, editors. Understanding and preventing HIV risk behavior: Safer sex and drug use. Thousand Oaks, CA: Sage Publications, Inc; 1996:80-99.
6. Bickham D, Moukalled S, Rich M. 232. Dating App Use and Sexual Risk Behaviors: Examining Aspects of Use and Motivation. Journal of Adolescent Health 2020 Feb;66(2):S117-S118. [doi: 10.1016/j.jadohealth.2019.11.235]
7. David G, Cambre C. Screened Intimacies: Tinder and the Swipe Logic. Social Media + Society 2016 Apr 07;2(2):205630511664197 [FREE Full text] [doi: 10.1177/2056305116641976]
8. Hammer J, Fisher J, Fitzgerald P, Fisher W. When Two Heads Aren't Better Than One: AIDS Risk Behavior in College-Age Couples 1. J Appl Social Pyschol 1996 Mar;26(5):375-397 [FREE Full text] [doi: 10.1111/j.1559-1816.1996.tb01855.x]
9. Huang ET, Williams H, Hocking JS, Lim MS. Safe Sex Messages Within Dating and Entertainment Smartphone Apps: A Review. JMIR Mhealth Uhealth 2016 Nov 08;4(4):e124 [FREE Full text] [doi: 10.2196/mhealth.5760] [Medline: 27826133]
10. Dijkstra P, Buunk BP, Blanton H. The Effect of Target's Physical Attractiveness and Dominance on STD-Risk Perceptions. J Appl Social Pyschol 2000 Aug 19;30(8):1738-1755 [FREE Full text] [doi: 10.1111/j.1559-1816.2000.tb02465.x]
11. Eleftheriou A, Bullock S, Graham CA, Stone N, Ingham R. Does attractiveness influence condom use intentions in heterosexual men? An experimental study. BMJ Open 2016 Jun 17;6(6):e010883 [FREE Full text] [doi: 10.1136/bmjopen-2015-010883] [Medline: 27315834]
12. Blanton H, Gerrard M. Effect of sexual motivation on men's risk perception for sexually transmitted disease: there must be 50 ways to justify a lover. Health Psychol 1997 Jul;16(4):374-379. [doi: 10.1037//0278-6133.16.4.374] [Medline: 9237090]
13. Fishbein M, Hennessy M, Yzer M, Curtis B. Romance and risk: romantic attraction and health risks in the process of relationship formation. Psychology, Health & Medicine 2004 Aug;9(3):273-285. [doi: 10.1080/13548500410001721846]

14. Dion K, Berscheid E, Walster E. What is beautiful is good. J Pers Soc Psychol 1972 Dec;24(3):285-290. [doi: 10.1037/h0033731] [Medline: 4655540]
15. Eagly AH, Ashmore RD, Makhijani MG, Longo LC. What is beautiful is good, but . . .: A meta-analytic review of research on the physical attractiveness stereotype. Psychological Bulletin 1991;110(1):109-128. [doi: 10.1037/0033-2909.110.1.109]
16. Kunda Z. The case for motivated reasoning. Psychol Bull 1990 Nov;108(3):480-498. [doi: 10.1037/0033-2909.108.3.480] [Medline: 2270237]
17. Chapman S, Hodgson J. Showers in raincoats: attitudinal barriers to condom use in high-risk heterosexuals. Community Health Stud 1988;12(1):97-105. [doi: 10.1111/j.1753-6405.1988.tb00576.x] [Medline: 3370950]
18. Jackson LA, Hunter JE, Hodge CN. Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review. Social Psychology Quarterly 1995 Jun;58(2):108. [doi: 10.2307/2787149]
19. Talamas SN, Mavor KI, Perrett DI. Blinded by Beauty: Attractiveness Bias and Accurate Perceptions of Academic Performance. PLoS One 2016;11(2):e0148284 [FREE Full text] [doi: 10.1371/journal.pone.0148284] [Medline: 26885976]
20. Feingold A. Good-looking people are not what we think. Psychological Bulletin 1992;111(2):304-341. [doi: 10.1037/0033-2909.111.2.304]
21. Cann A, Siegfried WD, Pearce L. Forced attention to specific applicant qualifications: Impact on physical attractiveness and sex of applicant biases. Personnel Psychology 1981 Mar;34(1):65-75. [doi: 10.1111/j.1744-6570.1981.tb02178.x]
22. Palmer CL, Peterson RD. Halo Effects and the Attractiveness Premium in Perceptions of Political Expertise. American Politics Research 2016;44(2):353-382. [doi: 10.1177/1532673x15600517]
23. Mulford M, Orbell J, Shatto C, Stockard J. Physical Attractiveness, Opportunity, and Success in Everyday Exchange. American Journal of Sociology 1998 May;103(6):1565-1592. [doi: 10.1086/231401]
24. Agocha VB, Cooper ML. Risk Perceptions and Safer-Sex Intentions: Does a Partner's Physical Attractiveness Undermine the Use of Risk-Relevant Information? Pers Soc Psychol Bull 1999;25(6):751-765. [doi: 10.1177/0146167299025006009]
25. Kahneman D. Thinking, fast and slow. New York: Macmillan; Oct 25, 2011.
26. Brown TA, Cash TF, Noles SW. Perceptions of Physical Attractiveness Among College Students: Selected Determinants and Methodological Matters. The Journal of Social Psychology 1986 Jun;126(3):305-316. [doi: 10.1080/00224545.1986.9713590]
27. Surawski MK, Ossoff EP. The effects of physical and vocal attractiveness on impression formation of politicians. Curr Psychol 2006 Mar;25(1):15-27. [doi: 10.1007/s12144-006-1013-5]
28. Camerer CF, Hogarth RM. The effects of financial incentives in experiments: A review and capital-labor-production framework. Journal of Risk and Uncertainty 1999;19:7-42. [doi: 10.1023/A:1007850605129]
29. Center for Open Science. URL: https://osf.io/gtcwd/ [accessed 2020-05-15]
30. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science 1974 Sep 27;185(4157):1124-1131. [doi: 10.1126/science.185.4157.1124] [Medline: 17835457]
31. Lichtenstein S, Fischhoff B. Do those who know more also know more about how much they know? Organizational Behavior and Human Performance 1977 Dec;20(2):159-183. [doi: 10.1016/0030-5073(77)90001-0]
32. Jankowiak WR, Hill EM, Donovan JM. The effects of sex and sexual orientation on attractiveness judgments: An evolutionary interpretation. Ethology and Sociobiology 1992 Mar;13(2):73-85. [doi: 10.1016/0162-3095(92)90019-z]
33. Strassberg DS, Lowe K. Volunteer bias in sexuality research. Arch Sex Behav 1995 Aug;24(4):369-382. [doi: 10.1007/BF01541853] [Medline: 7661653]
34. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. Psychol Bull 1990 Nov;108(3):339-362. [doi: 10.1037/0033-2909.108.3.339] [Medline: 2270232]
35. Wiederman M. Volunteer bias in sexuality research using college student participants. Journal of Sex Research 1999 Feb;36(1):59-66 [FREE Full text] [doi: 10.1080/00224499909551968]
36. Eisenberg ML, Shindel AW, Smith JF, Lue TF, Walsh TJ. Who is the 40-year-old virgin and where did he/she come from? Data from the National Survey of Family Growth. J Sex Med 2009 Aug;6(8):2154-2161. [doi: 10.1111/j.1743-6109.2009.01327.x] [Medline: 19493289]
37. Olderbak SG, Malter F, Wolf PS, Jones DN, Figueredo AJ. Predicting Romantic Interest at Zero Acquaintance: Evidence of Sex Differences in Trait Perception but Not in Predictors of Interest. Eur J Pers 2017;31(1):42-62 [FREE Full text] [doi: 10.1002/per.2087] [Medline: 28736483]
38. Khan KS, Chaudhry S. An evidence-based approach to an ancient pursuit: systematic review on converting online contact into a first date. Evid Based Med 2015 Apr;20(2):48-56. [doi: 10.1136/ebmed-2014-110101] [Medline: 25678447]
39. McFarlane M, Bull SS, Rietmeijer CA. The Internet as a newly emerging risk environment for sexually transmitted diseases. JAMA 2000 Jul 26;284(4):443-446. [doi: 10.1001/jama.284.4.443] [Medline: 10904506]
40. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]
41. Askelson NM, Campo S, Smith S, Lowe JB, Dennis L, Andsager J. Assessing physicians' intentions to talk about sex when they vaccinate nine-year-old to 15-year-old girls against HPV. Sex Education 2011 Aug 12;11(4):431-441. [doi: 10.1080/14681811.2011.595252]

42. Kuehn BM. A Proactive Approach Needed to Combat Rising STIs. JAMA 2019 Jan 29;321(4):330-332. [doi: 10.1001/jama.2018.20400] [Medline: 30624545]

Abbreviations

IRB: institutional review board
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 02.04.19; peer-reviewed by L Watchirs Smith, G Sabben, KA Nguyen; comments to author 30.10.19; revised version received 04.02.20; accepted 29.03.20; published 09.06.20.

Please cite as:
Krishnamurti T, Davis AL, Fischhoff B
Inferring Sexually Transmitted Infection Risk From Attractiveness in Online Dating Among Adolescents and Young Adults: Exploratory Study
J Med Internet Res 2020;22(6):e14242
URL: https://www.jmir.org/2020/6/e14242
doi:10.2196/14242
PMID:32442128

Original Paper

Toward the Development of Data Governance Standards for Using Clinical Free-Text Data in Health Research: Position Paper

Kerina H Jones[1], BSc, PhD; Elizabeth M Ford[2], MA, DPhil; Nathan Lea[3], BA, MSc, PhD; Lucy J Griffiths[1], BSc, MSc, PhD; Lamiece Hassan[4], BSc, PhD; Sharon Heys[1], BSc; Emma Squires[1], BSc; Goran Nenadic[5], BSc, MSc, PhD

[1]Population Data Science, Medical School, Swansea University, Swansea, United Kingdom
[2]Brighton and Sussex Medical School, Brighton, United Kingdom
[3]Institute of Health Informatics, University College London, London, United Kingdom
[4]Division of Informatics, Imaging & Data Sciences, University of Manchester, Manchester, United Kingdom
[5]Department of Computer Science, University of Manchester & The Alan Turing Institute, Manchester, United Kingdom

Corresponding Author:
Kerina H Jones, BSc, PhD
Population Data Science
Medical School
Swansea University
Data Science Building
Singleton Park
Swansea, SA28PP
United Kingdom
Phone: 44 01792602764
Email: k.h.jones@swansea.ac.uk

Abstract

Background: Clinical free-text data (eg, outpatient letters or nursing notes) represent a vast, untapped source of rich information that, if more accessible for research, would clarify and supplement information coded in structured data fields. Data usually need to be deidentified or anonymized before they can be reused for research, but there is a lack of established guidelines to govern effective deidentification and use of free-text information and avoid damaging data utility as a by-product.

Objective: This study aimed to develop recommendations for the creation of data governance standards to integrate with existing frameworks for personal data use, to enable free-text data to be used safely for research for patient and public benefit.

Methods: We outlined data protection legislation and regulations relating to the United Kingdom for context and conducted a rapid literature review and UK-based case studies to explore data governance models used in working with free-text data. We also engaged with stakeholders, including text-mining researchers and the general public, to explore perceived barriers and solutions in working with clinical free-text.

Results: We proposed a set of recommendations, including the need for authoritative guidance on data governance for the reuse of free-text data, to ensure public transparency in data flows and uses, to treat deidentified free-text data as potentially identifiable with use limited to accredited data safe havens, and to commit to a culture of continuous improvement to understand the relationships between the efficacy of deidentification and reidentification risks, so this can be communicated to all stakeholders.

Conclusions: By drawing together the findings of a combination of activities, we present a position paper to contribute to the development of data governance standards for the reuse of clinical free-text data for secondary purposes. While working in accordance with existing data governance frameworks, there is a need for further work to take forward the recommendations we have proposed, with commitment and investment, to assure and expand the safe reuse of clinical free-text data for public benefit.

(J Med Internet Res 2020;22(6):e16760) doi:10.2196/16760

KEYWORDS

ethical; legal; social implications; public engagement; free-text data; information governance

Introduction

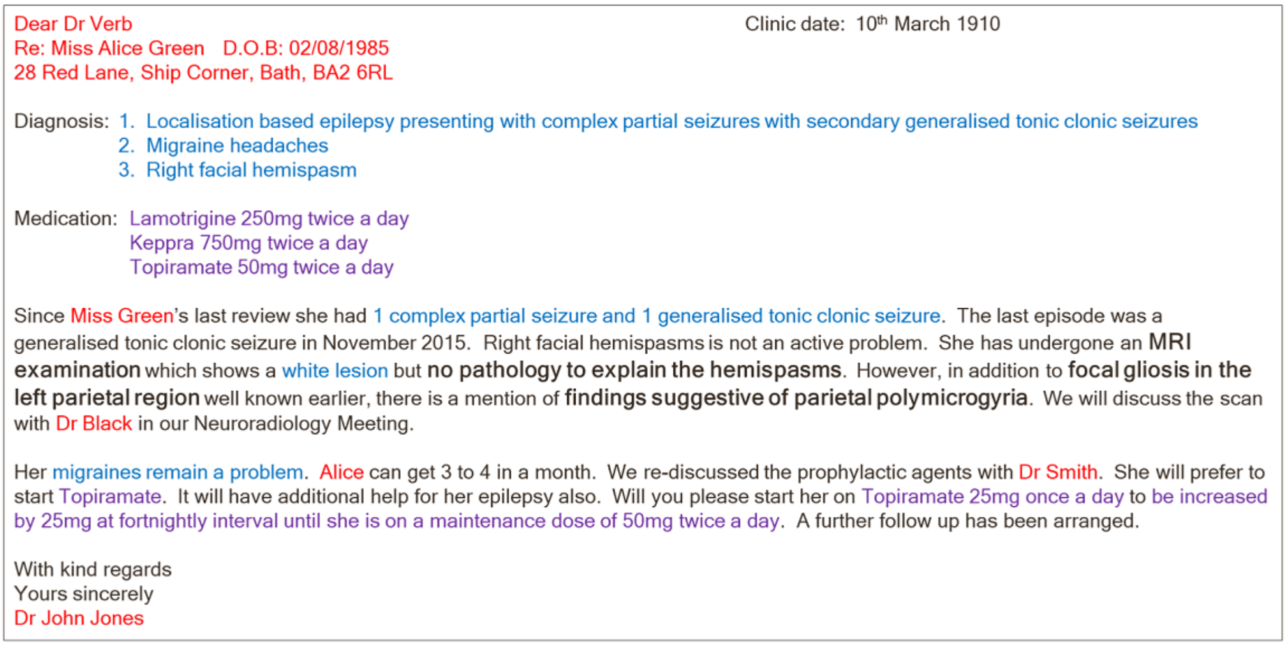
Background

Structured electronic health records (EHRs) have long been used in large-scale research to create new knowledge to inform clinical care, practice, and policy. There are many enterprises specializing in making EHR data available in accordance with jurisdictional legislation and governance [1-4]. We refer to structured data as information recorded in specified forms and fields within the EHRs. Generally, clinical data are highly personal and sensitive and, therefore, need to be deidentified or anonymized before they can be used for any secondary purposes outside the clinical environment. We refer to deidentified data as records from which the commonly recognized identifiers (eg, name, address, and date of birth) have been removed and anonymized data as information from which the data subject cannot be identified [5]. EHR data also contain free-text components, including notes made at consultations and referral letters, which, by definition, are unstructured. As such, they tend to be less available beyond the immediate setting and are typically not readily available for secondary uses such as research or service improvement. However, clinical free-text data represent a vast, untapped source of rich information to guide research and clinical care, including patient-specific context and details that clarify and supplement information coded in structured data fields. Furthermore, some clinical information in mental health, pathology, and imaging reports is not available in coding structures but only in free-text form. By their nature, free-text data are highly likely to contain sensitive information and patient identifiers, and this must be suitably processed to protect individual privacy before the information can be shared further. The key difficulty is that there are major challenges in finding effective methods to deidentify free-text on a scale that does not damage data utility as a by-product.

Methods from natural language processing (NLP) have been developed to automatically scan clinical free-text data to identify potential identifiers, such as patients’ names, significant dates, and their family members. There are generally 2 approaches to automated deidentification. In a method referred to as *blacklisting*, deidentification algorithms remove identifying information so that these variables are masked in the data extract. Alternative methods focus only on isolating the relevant clinical information from personal identifiers via extraction of specified variables such as medication dosage instructions or diagnoses, which are *whitelisted* and preserved in text. Whitelisting can be thought of as the converse of blacklisting in that it extracts clinically informative data rather than excluding disallowed pieces of information [6-8].

Each method has advantages and disadvantages and many permutations within them. For example, blacklisting can result in some personal identifiers remaining within deidentified text or important nonidentifiable data being removed from the text if the process is too stringent. Conversely, whitelisting can result in valuable data being left behind or false information being extracted inadvertently. Whitelisting also relies on a data extraction algorithm to choose the relevant clinical information, if no human is to see the identifiable data. A synthetic clinic letter is shown in Figure 1 to illustrate some of these challenges. As can be seen, a clinic letter can contain a variety of identifiable information about the patient and their care providers. Accuracy and the resulting privacy and utility in preparing free-text extracts for reuse depends on the sophistication of the methodologies in use. We provide some simple example problems for illustration purposes. Blacklisting could result in some personal identifiers remaining within purportedly deidentified text, for example, names that are also common words—Green and Verb. Conversely, whitelisting can leave valuable data behind, for example, *parietal polymicrogyria* without the word *suggestive* or *pathology to explain the hemispasms* without extracting the word *no*. There is much natural language processing and other methodological work underway to enable the creation of extracts that retain data utility and protect privacy. It is important to note that deidentification and extraction algorithms do not work *out of the box* but often have to be built and tested on specific data annotated by domain-specific experts to train and develop the algorithms. In general, algorithms can be trained to work to a standard comparable with that of a human annotator, but accuracy can decrease with increasing information complexity [6-8].

Figure 1. Synthetic clinic letter.



Dear Dr Verb
Re: Miss Alice Green D.O.B: 02/08/1985
28 Red Lane, Ship Corner, Bath, BA2 6RL

Clinic date: 10th March 1910

Diagnosis:
1. Localisation based epilepsy presenting with complex partial seizures with secondary generalised tonic clonic seizures
2. Migraine headaches
3. Right facial hemispasm

Medication: Lamotrigine 250mg twice a day
Keppra 750mg twice a day
Topiramate 50mg twice a day

Since Miss Green's last review she had 1 complex partial seizure and 1 generalised tonic clonic seizure. The last episode was a generalised tonic clonic seizure in November 2015. Right facial hemispasms is not an active problem. She has undergone an **MRI examination** which shows a white lesion but **no pathology to explain the hemispasms**. However, in addition to **focal gliosis in the left parietal region** well known earlier, there is a mention of **findings suggestive of parietal polymicrogyria**. We will discuss the scan with Dr Black in our Neuroradiology Meeting.

Her migraines remain a problem. Alice can get 3 to 4 in a month. We re-discussed the prophylactic agents with Dr Smith. She will prefer to start Topiramate. It will have additional help for her epilepsy also. Will you please start her on Topiramate 25mg once a day to be increased by 25mg at fortnightly interval until she is on a maintenance dose of 50mg twice a day. A further follow up has been arranged.

With kind regards
Yours sincerely
Dr John Jones

Being able to create free-text extracts that do not contain identifying information with a high degree of confidence is a fundamental issue but is not the only consideration for the governance of clinical free-text data. In addition, there are issues such as regulatory approvals including independent ethics committee review, working with or within the health service, accessing identifiable data for the deidentification and extraction process, engaging with clinicians to understand clinical concerns and point of care issues, patient involvement, questions around patient consent, retaining the ability for data linkage, and conditions and environments for data reuse.

Main Aim

In recognition of these challenges, the main aim of this study was to develop recommendations for the creation of data governance standards to integrate with existing frameworks for personal data use, to enable free-text data to be used safely for research for patient and public benefit. As such, it will form a position paper and inform further work. In this context, we define *safely* as where the identity of the individual is highly unlikely to be discovered. We refer to this study as TexGov. It focuses on data governance and health services within the United Kingdom, but the main findings are applicable more widely.

Methods

Study Design

TexGov was designed to cover a variety of activities to develop data governance standards for using clinical free-text data in research. It included an outline of the UK data protection landscape for context, a rapid literature review on governance aspects in previous UK research studies using clinical free-text data, UK case studies with systems providing access to free-text data for more in-depth information, engagement with researchers to explore barriers and solutions in working with free-text data, and engagement with the public for views on socially acceptable approaches. These activities are described in the following sections.

Data Protection Landscape

We reviewed the relevant legislation, regulations, and official guidance in place in the United Kingdom for context on governance aspects for the use of free-text data. This included the UK Data Protection Act (DPA) 2018 [9], the European Union General Data Protection Regulation (GDPR) 2016 [5], the Human Rights Act [10], the common law duty of confidentiality (CLDC [11]), the Caldicott Principles [11], and the Information Commissioner's Office guidance on data sharing and anonymization [12,13].

Free-Text Data Governance Practice

We gained information on data governance in working with clinical free-text data by carrying out a rapid literature review [14] and mini-case studies of systems making deidentified free-text data available for research in the United Kingdom. The methodology for the literature review is given below:

Search Strategy

Searches for studies indexed in PubMed and Web of Science (WoS) were conducted on January 15, 2019, using the following search terms: (1) "electronic health records" or "electronic medical records" or "electronic patient records" or "hospital records" or "personal health records" or "computerized patient records" or "computerized medical records" combined with (2) "text mining" or "natural language processing" or "free text" or "narrative." Given our sole interest in ethical and governance procedures in the United Kingdom, the search was then restricted to papers using UK databases containing free-text data by using the following terms: "UK" or "United Kingdom" or "Britain" or "England" or "Wales" or "Ireland" or "Scotland." The search was limited to human studies published in English. No restrictions were imposed on the year of publication.

Eligibility

To be eligible for this review, published research had to meet the following criteria: primary research using free-text records; UK-based health databases or data from UK hospitals; and information extracted from the text of (human, not veterinary) electronic medical records, medical letters or medical reports, and methods papers (eg, development or description of search and analytic tools), but only if the paper used patient data in the study.

Data Extraction

Characteristics of the included studies were extracted by one researcher and reviewed by another researcher and were as follows: (1) year published and lead author institution, (2) broad purpose of the study, (3) what and where data were accessed, (4) focus of free-text data, (5) whether the data were accessed in identifiable form, (6) whether the study used the health records in conjunction with other data, (7) main finding or conclusion, and (8) any ethics and governance detailed. We will focus on data governance aspects only, rather than the clinical value of using free-text data.

Case Studies

The case studies were conducted through face-to-face or telephone interviews in February 2019 and were structured to capture information about the main model for making data available, types of data included, data linkage capability, free-text deidentification or extraction method, and governance approvals. Notes were taken during the interviews and subsequently checked for accuracy with each interviewee.

Engagement With Clinical Text-Mining Researchers

We held a 1-day workshop (January 16, 2019) for clinical free-text mining researchers to explore their perceptions of barriers and solutions in working with free-text data for sharing outside the clinical setting. Participants were recruited via advertisements on the Healtex website. The workshop was attended by 44 people and included outlines of the TexGov study aims, the findings of a previous citizens' jury on using clinical free-text data for research [15], and data protection. These were followed by presentations from the NLP community on their approaches to working with clinical free-text data, during which the audience was asked to write down the topics to discuss in more detail. The identified topics were grouped into 4 themes and used in group discussions on challenges and solutions: (1) patient involvement at identifiable and deidentified data stages, (2) opt-in/opt-out consent models for the reuse of free-text data, (3) working with identifiable data for NLP algorithm development, and (4) deidentification methods and thresholds of reliability. Delegates were randomly allocated to one of the groups but were allowed to change to another if they felt strongly about a particular topic. Each of the discussions was facilitated and noted to capture views on the nature of the challenge, what can be done to address it, and how that can be achieved.

Engagement With the Public

We discussed the TexGov study with 2 public groups. The first event was advertised to the general public and patient interest groups by the Alan Turing Institute and held at their premises on March 28, 2019, attended by approximately 50 people. It included a presentation on the TexGov study and a role-play exercise to illustrate how clinical free-text data are collected at a hypothetical general practitioner (GP) consultation and to engage the audience in discussion. It also included a panel discussion so the audience could ask questions to the representatives from the National Data Guardian's office [16]; Understanding Patient Data, an independent patient data taskforce [17]; use MY Data, a movement of patients, carers, and relatives in support of using patient data [18]; and National Health Service (NHS) England [19]. Notes were taken during the panel discussion. This was followed by facilitated and noted group discussions on the topics of (1) Transparency and patient choice in the use of free-text data: what information do patients need and which are the best methods of dissemination? (2) Identifiability, deidentification, and anonymization: what is the best approach to making data available? and (3) Data access models and security in working with free-text data: how to balance restricting and facilitating access to data? These topics were prioritized from issues for discussion identified by registered delegates in advance of the workshop.

The second public engagement activity was conducted with the consumer panel based at Swansea University on May 1, 2019. This panel is made up of 18 members of the general public, and it advises on developments in the use of person-based data for research, such as via the Secure Anonymised Information Linkage (SAIL) Databank [20]. Following a description of the TexGov study and the preliminary findings, there was an open discussion so that members of the panel could provide their views on the work and the publicly acceptable way forward for the reuse of free-text data. Notes were taken on the views of the panel members.

Results

Data Protection Landscape

The Human Rights Act provides overarching legislation, with Article 8 setting out rights to enjoy a private life free of intrusion and interference, subject to restrictions in accordance with law in a democratic society [10]. Our review of UK statutory data protection law in the DPA [9] and the GDPR [5] showed that free-text data are not singled out by these instruments: the provisions apply to free-text data in the same way as any general person-identifiable data (PID) and special category health data. Similarly, free-text data given in confidence, such as in a physician-patient consultation, are subject to the same CLDC [11] principles and professional practice as for structured, coded data. The Caldicott Principles [11] set out fundamental good practices for the protection of information that could identify an individual but also highlight that data sharing can be as important as the duty to protect confidentiality. Within these principles and in the Information Commissioner's Office guidance on data sharing and anonymization, there is nothing specifically stated about free-text data (Table 1). The requirements are for data use to be subject to an appropriate lawful basis and for data to be processed to protect privacy before secondary uses, unless another justification (such as participant consent for research) is in place. This is a challenge for the reuse of clinical free-text data: the difficulty in being confident (and providing evidence) that deidentification is adequate when making claims of anonymity or minimizing the use of personal data, so that information derived from the unstructured format of free-text data can be safely taken forward for reuse.

Table 1. Summary of relevant data protection landscape.

Item	Key points
UK Human Rights Act	Sets out an individual’s rights to enjoy a private life free of intrusion and interference
European Union General Data Protection Regulation and the UK Data Protection Act	Provisions for processing general person-identifiable data and special category health data
Common Law Duty of Confidence	Governs the use of data given in confidence, such as in a physician-patient consultation
Caldicott Principles	Set out good practice for the protection of information that could identify an individual and the importance of data sharing
Information Commissioner’s Office	Provides guidance on data sharing and anonymization

Data Governance

The search identified 45 papers from PubMed and 340 papers from WoS, with 354 papers remaining after the removal of duplicates. After screening abstracts and full texts, 51 studies were defined as eligible for inclusion in the rapid review. Medical records were accessed via 3 main databases: (1) the Health Improvement Network (THIN) primary care database [21,22] (N=7) [23-29], (2) South London and Maudsley NHS Foundation Trust (SLaM) with clinical records accessed via the Clinical Records Interactive Search (CRIS) system [30-32] (N=16) [33-49], and (3) the Clinical Practice Research Datalink (CPRD) [50] (N=13) [51-62]. One study used the SAIL Databank [63]. There were 13 independent studies, which were not associated with a data management infrastructure [64-76].

The THIN and CPRD databases contain NHS patient primary care records, whereas the CRIS system includes clinical records for people who have used a range of NHS mental health services and substance misuse services in the United Kingdom. The study that used the SAIL Databank accessed the primary care dataset. The independent studies accessed a range of free-text records, including from accident and emergency department, outpatient department, intensive care, primary care, and prescribing datasets. All records in the studies that used THIN, CRIS, CPRD, and SAIL Databank were anonymized, and no personal information was, therefore, available to researchers. The independent studies were more variable depending on specific approvals and permissions, for example, identifiable data accessed by authorized hospital staff within an NHS organization [64,71] or data extracted following the ethical approval of the study [76].

All studies reporting the use of THIN noted ethical approvals, and 3 studies also cited the overarching THIN data resource approvals by the NHS South East Multicenter Research Ethics Committee. All but one [48] of the studies that used the CRIS data resource noted approval by the Oxfordshire research ethics committee and the service user-led oversight committee study approval, which must be granted before access to the anonymized data is permitted. Of the 13 studies using CPRD free-text data, 4 provided no ethical approval details [51-53,61]. All the remaining studies in this group detail some level of ethical approval, either study approval or the resource multicenter research ethics committee approval for all observational research using CPRD data. The study accessing free-text data in SAIL Databank does not cite the information governance review panel (IGRP) approval that must have been sought before access to SAIL data was granted [63]. Of the 13 independent studies, 7 detailed the local ethical approval process [67-69,72,73,75,76]. One stated that ethical approval was not necessary, as the purpose of the study was to measure service provision retrospectively without using identifiable patient information [74], and the remaining 7 studies did not mention ethical approvals [64-66,70,71].

None of the 7 studies that used THIN detailed the data access model, which could have been via a sublicense agreement enabling access for a defined period to conduct unlimited studies, subject to protocol approval, or release of data extracts consisting of subsets of raw data in accordance with researchers’ study protocols and specifications. CRIS is accessed within the NHS system by researchers at the Maudsley Biomedical Research Centre; as such, most lead authors (or at least one coauthor) of the 16 studies using SLaM clinical records were based at King’s College Hospital. Some, but not all, studies specify this access route. CPRD (formally known and referenced in studies as the General Practice Research Database) operated on a data release model under the strict CPRD and data control terms. Few of the retrieved studies specified this model. The study using free-text records in the SAIL Databank did not detail the access route, that is, following necessary governance approvals and safe researcher training, data can be accessed remotely within the safe haven with no external data release [77]. The 13 independent studies provided very little information on how data were accessed. A summary of key information from the rapid review is given in Table 2.

Table 2. Key information from the literature review.

Institute	Study	Free-text search	Linkage to additional datasets	Use of identifiable data	Ethical approval
THIN[a] (n=7)	[23-29]	[23] cause of death; [24] ischemic cerebrovascular events; [25] diabetic retinopathy and diabetic maculopathy diagnoses; [26] child maltreatment; [27] colorectal cancer recording; [28] endometriosis diagnosis; and [29] major malformations and comments assigned to referrals to specialists	Yes: [23,25,27,28]; no: [24,26,29]	Yes: [26]; all others: no	THIN data resource approved by NHS South East Multicenter Research Ethics Committee
CRIS[b] (n=17)	[33-49]	[33] smoking status; [34] extrapyramidal side effects and adverse drug events; [35] antipsychotic polypharmacy; [36] negative symptoms of schizophrenia; [37] mood instability; [38] cannabis use; [39] medication descriptions; [40] characteristics of people with Alzheimer disease; [41] delivery of cognitive behavioral therapy for psychosis; [42] notes and correspondence on diagnoses of mental disorders; [43] adverse drug events; [44] symptoms of severe mental illness; [45] risk factors for depression; [46] presence of negative symptoms and antipsychotic use; [47] suicide ideation and attempts; [48] information on hepatitis C and HIV; and [49] registered company addresses	Yes: [39,42,45]	No	CRIS approved for secondary analysis by the Oxfordshire Research Ethics Committee. A service user-led oversight committee considers all proposals before access to the anonymized data is permitted
CPRD[c] (n=13)	[51-62]	[51] heart defects; [52] congenital malformations; [53] pregnancy outcomes; [54] ovarian cancer diagnoses; [55] cause of death; [56] coronary angiogram results; [57] keywords for rheumatoid arthritis; [58] markers for rheumatoid arthritis; [59] disease-modifying antirheumatic drugs; [60] records of visible hematuria, jaundice, or abdominal pain; [61] drug usage values and administrations; and [62] terms indicating allergic bronchopulmonary aspergillosis cases	Yes: [51-53]; no: [54-62]	No	Multicenter research ethics committee approval was in place for all observational research using CPRD data
SAIL[d] (n=1)	[63]	[63] symptoms of ankylosing spondylitis	No	No	Ethical approval is not required for the use of anonymized data within SAIL. An independent information governance review panel assesses all proposed uses of SAIL data.

Institute	Study	Free-text search	Linkage to additional datasets	Use of identifiable data	Ethical approval
Independent studies (n=13)	[64-76]	[64] string "asth" for asthma; [65] range of clinical terms; [66] reason for admission; [67] test results for stages 3-5 chronic kidney disease; [68] search in reports of CT[e] scan of brain for stroke; [69] search in reports of CT scan of brain for stroke, subarachnoid hemorrhage, or ischemic stroke; [70] keywords on hearing aid decisions; [71] reasons for dose omissions; [72] reasons for deaths due to unsafe care; [73] breathlessness and wheeze symptoms; [74] focal liver lesions; [75] cardiology information; and [76] clinician discourses compared with patient narratives	Yes: [68,69,76]; no: [64-67], [70-75]	Yes: [64,71,73,76]; no: [67-69,72,74,75]; and not clearly stated: [65,66,70]	Variable depending on particular study

[a] THIN: The Health Improvement Network.

[b] CRIS: Clinical Records Interactive Search.

[c] CPRD: Clinical Practice Research Datalink.

[d] SAIL: Secure Anonymised Information Linkage.

[e] CT: computed tomography.

Table 2 summarizes the key information reported in the studies included in the rapid literature review. The first column shows the number of studies relating to each institute. The study numbers in subsequent columns are as per the reference list (and are not counts of studies). The extent to which individual studies reported approvals was variable, although (at least for those attached to THIN, CRIS, CPRD, or SAIL) they would have had to abide by the mechanisms in place before data access was granted because access is under institutional control. Rather than attempting to include all details by study, we showed the standard approvals in the final column with fuller information in the text.

The interviews provided further information on systems providing access to data derived from clinical free-text sources. The purpose was to learn about their methods and from their experiences. We do not describe the models in full detail, as that would be beyond the scope of this paper. The THIN database was created in 2003 by In Practice Systems (Vision). THIN is a primary care data repository with the capability to extract free-text as well as structured data from GP practices, but currently, only structured data are released to external researchers. THIN has collaborated with University College London (UCL) to promote academic use of the data. As part of this, a copy of the THIN database will sit in a UCL data safe haven, which can be accessed by UCL researchers. THIN is seeking ethical approval to enable the use of the free-text data for academic research within UK academic or NHS safe havens. Free-text data will be subjected to an automated deidentification process using a blacklisting method. The full data governance model at UCL is in the final stages of development at the time of writing. THIN is overseen by an advisory committee involving patient, clinician, and researcher representatives, and all studies are reviewed by a scientific review committee before data access can be granted [21,22].

CRIS is a medical record inquiry program that has been implemented in a number of mental health trusts. Our example relates to SLaM, and so, we refer to *CRIS at SLaM* because other NHS Trusts might not operate CRIS under exactly the same principles. Data are held in a repository inside the Trust firewall (ie, in the same domain as the original health record), which comprises a database of structured and free-text mental health records that are deidentified, but linkable at the individual level, before entering the repository. Unlike some blacklisting algorithms that rely on a lexicon to ascertain and remove names, the CRIS deidentification algorithm has access to individual health records and makes reference to this to blank out the name of the patient. CRIS at SLaM has ethical and s251 approval: the latter for occasions when PID need to be exported outside the firewall to enable linkage. It also has approval from the Trust Caldicott Guardian and the executive board. The person accessing the data in the safe haven must have a SLaM contract, a SLaM honorary contract, or a research passport (an accreditation issued by the Health Research Authority [78]). All projects must include at least one member of the study team who has an honorary or substantive SLaM contract and can act as a guarantor [30-32].

SAIL is a repository that holds multiple health and wider administrative datasets in a deidentified form about the population of Wales. SAIL Databank does not process PID but uses a trusted third party to receive the PID and carry out a matching and deidentification process, with the creation of a consistent identifier unique for each person represented to enable individual-level linkage across datasets. Further controls are enacted to provide access to data for research in anonymized form within a safe haven. The SAIL data comprise extracts of structured, coded data, with minimal free-text data at present. However, work is underway to incorporate free-text data so that they can be included in studies when needed. Unlike approaches that deidentify free-text data (blacklist) and incorporate the remainder, the algorithm used by SAIL extracts medical terms and descriptors (whitelist) at the NHS source for incorporation so that only nonidentifiable structured data leave the organization. SAIL Databank took this approach because of the risks of introducing insufficiently deidentified information into the databank. Working with free-text data at source is by means of an NHS honorary contract [8], and all proposals to use SAIL data must have received approval from an independent IGRP before access can be granted via the data safe haven [77].

CPRD operates on a data release model and used to collect and release deidentified (blacklisted) free-text data under strictly controlled protocols. However, this was discontinued in 2013 on advice from the Information Commissioner's Office, and the catalog of free-text data was destroyed. At present, CPRD has no plans to develop NLP or an automated text anonymization service. Instead, they liaise directly with GPs on behalf of researchers to validate codes to supplement coded data with additional information.

Engagement With Clinical Text-Mining Researchers

A summary of the information gained during the discussions on each theme is shown below in relation to the nature of the challenge, what could be done to address it, and how this could be achieved.

Patient Involvement at Identifiable and Deidentified Data Stage

This theme relates to the principles of engagement with the public for good research practice in general.

Nature of the Challenge

Delegates felt that there was a challenge in balancing what could be perceived as a lack of individual choice in nonconsented data cohorts with the risk of selection bias in consented cohorts, particularly if there is doubt about the level of knowledge among individuals, including conceptual differences between coded and free-text data.

What Can Be Done

Additional public engagement in free-text research was seen as needed. This should be linked with wider information about patient data uses to help people better understand the issues and gain their input. This could contribute to a checklist of recommendations for researchers using free-text data, so the public views on the use of free-text data are taken into account.

How It Can Be Achieved

Public panels and existing groups can help clarify reasonable expectations regarding the use of free-text data. It would be useful to have a repository of findings from research using free-text data showing its clinical benefits. This could also include information in lay format on advances in methodological research to improve processes of handling free-text data.

Opt-In/Opt-Out Consent Models for Reuse of Free-Text Data

This theme relates to managing individual choices in the use of free-text data for research.

Nature of the Challenge

Discussions focused on the need to convey a reasonable understanding of the motivations, risks, and benefits of researchers accessing free-text data and how data reuse is governed. In particular, the challenge of providing appropriate consent options was highlighted.

What Can Be Done

As consent might not be required, or even be possible, for processing large-scale free-text data across populations, delegates stressed the need for guidance and good practice on free-text data reuse in line with governance requirements and expectations.

How It Can Be Achieved

A wider examination of existing practice used by government departments along with clinical and regulatory codes of practice would provide further guidance on how to manage consent options. Possible solutions could be piloted and evaluated for practicality.

Working With Identifiable Data for Natural Language Processing Algorithm Development

This theme relates to the practicalities in accessing free-text data in identifiable form to prepare extracts for research.

Nature of the Challenge

Free-text data are needed in their original (raw) form for effective NLP algorithm development and maintenance, yet the data are identifiable, and some contain sensitive information. In addition, the free-text datasets accessed need to be of sufficient scale to avoid introducing bias when developing NLP methods.

What Can Be Done

There is a need to clarify whether enabling data access via a safe haven can provide a suitable solution, if patient data can be used without consent for the purpose of training algorithms, and the usefulness of publicly available datasets.

How It Can Be Achieved

Delegates suggested possible solutions by creating a *Text Bank* to support NLP development via data donation, the use of synthetic datasets, and sharing annotated free-text data with other developers.

Deidentification Methods and Reliability Thresholds

This theme relates to residual reidentification risk and data utility in the datasets used for research.

Nature of the Challenge

Delegates focused on the challenges in quantifying the residual risk of reidentification in free-text data extracts, particularly in the face of the high degree of data variability and the absence of a gold standard threshold for free-text deidentification.

What Can Be Done

Discussions highlighted the need for robust security with only trusted access and an audit trail, and that free-text data extracts should be treated as potentially identifiable. More research on the relationship between the accuracy of the deidentification algorithm, risk of reidentification, and data utility was identified as a need.

How It Can Be Achieved

Accessing free-text data extracts only within safe havens by small teams embedded within organizations, and with clear lines of accountability, was seen as a positive step. A range of risk models should be employed to avoid overscrubbing the data with consequent loss of data utility and to determine acceptable levels of deidentification.

Engagement With the Public: Panel Discussions at the Turing Institute Event

The panel discussion at the Turing Institute event with representatives from the National Data Guardian’s office, Understanding Patient Data, use MY Data, and NHS England raised some important points. The National Data Guardian representative highlighted that data protection legislation is not a barrier to data sharing, but there is a need for proper transparency and trustworthiness to build public confidence. In addition, there is a need for a joined-up approach in developing consistent standards, language, and message. The representative from understanding patient data stressed the huge potential in being able to use clinical free-text data beyond direct care but also highlighted the challenges in being able to enact this in an acceptable way, respecting privacy and ensuring accuracy in data extracted. They also spoke of the need for evidence of the potential benefits to patients of the use of their clinical free-text data. The representative from use MY Data emphasized the enormous potential of, as yet, largely underutilized free-text data for research purposes, seeing this as an *atrocious waste*. This was accompanied by an expression of the need for ongoing patient engagement and that use MY Data is an instruction, not a request, as the group wants greater use of patient data and is working hard to convey this message to decision makers. The NHS England representative spoke about the CLDC and the need for free-text data use not to impact the confidentiality of medical consultations. All the panelists emphasized the need to demonstrate to the public the benefits of using clinical free-text data and the importance of working in accordance with existing research and information governance frameworks, but that adaptations to these frameworks may be required for emerging data types and formats.

Engagement With the Public: Group Discussions at the Turing Institute Event

A summary of the points emerging from the group discussions with delegates at the Turing Institute event is provided in relation to knowledge gaps, involvement and engagement, and suggested solutions.

Transparency and Patient Choice in the Use of Clinical Free-Text Data: What Information Do Patients Need and Which Are the Best Methods of Dissemination?

Knowledge Gaps

The delegates acknowledged the efforts that go into public involvement and engagement but highlighted the seldom heard voices, such as hard to reach groups, marginalized groups, and young people.

Involvement and Engagement

They felt that because of risks of misunderstanding, information for the public should be layered, that is, it should be contextualized by founding it on health data and research before adding the particular features, risks, and benefits of free-text data.

Suggested Solutions

Information should be provided in plain language and in accessible formats, taking into account differing needs and abilities. Delegates recommended tapping into patient networks as authentic, credible ambassadors for discussions with the wider public and the value in working with communication experts. It was seen as important to explain that although there are challenges in effectively deidentifying free-text data, it is a format of data and not fundamentally different from other health data.

Identifiability, Deidentification, and Anonymization: What Is the Best Approach to Making Data Available?

Knowledge Gaps

The delegates felt unclear about how the GDPR (introduced in 2018) would apply within the context of the existing UK regulations and uncertainties around how best to make data available and minimize risks. There were uncertainties around how one might seek informed consent to participate in research in a way that participants would understand but also how wider public would be able to grasp concepts in the abstract.

Involvement and Engagement

The delegates expressed the need for clearer information on the relationship between having a lawful basis for data processing under the GDPR and the fact that this does not negate the need for consent to participate in research. With regard to identifiability, delegates felt that the research community should make a strong case for the use of free-text data in research, including that it was important not to lose that richness.

Suggested Solutions

Delegates felt that engagement and educational events should be carefully planned in terms of how they could empower people to make informed decisions when it came to reusing of free-text data. This would need to provide a compelling reason for research to look at minimally deidentified free-text data to preserve the richness, balanced against how the risks to participants could be managed.

Data Access Models and Security in Working With Free-Text Data: How to Balance Restricting and Facilitating Access to Data?

Knowledge Gaps

Delegates acknowledged a general lack of awareness of different data access models and the degrees to which data can be controlled, shared, or made openly accessible. They stressed the risks of not using the free-text data and the benefits that could be added to research.

Involvement and Engagement

Delegates felt that there should be more involvement and engagement with the public to explain due diligence processes and safeguards in managing free-text data. This should include interactive methods and demonstrations, not limited to passive information transfer.

Suggested Solutions

Delegates stressed that free-text data need to be used so that the United Kingdom does not lag behind other countries. The research community should be open to the public and explain that deidentification of free-text data may be imperfect. Patient groups should follow the example of use MY Data and lobbying data controllers for free-text data to be made more accessible to researchers.

Engagement With the Public: Consumer Panel

Having been presented with the findings of the previous engagement activities (described here), the consumer panel agreed and furthermore suggested that patient involvement should be included at all stages practical in the development of algorithms to avoid bias in relation to diversity; that involvement should be ongoing because of societal attitudinal changes over time and as new knowledge comes to light; and that a databank of clinical free-text data for algorithm development would be a good thing, provided that it is done transparently and properly managed.

Discussion

Principal Findings

To our knowledge, this is the first study to explore the data governance aspects for using clinical free-text data and to do this through a combination of outlining the legislative and regulatory backdrop, reviewing literature and systems using free-text data extracts in health research, and gaining the perspectives of free-text data researchers and members of the public. In terms of UK data protection legislation and guidance, there is nothing specific in relation to free-text data. In many ways, this was to be expected, as legislation tends to be high level needing interpretation and justification of an appropriate lawful basis on how data are to be used. In addition, free-text is a data format and not a different type or category (eg, health, political views, and ethnicity) of data, and so, it is covered by the data protection provisions for health data. In terms of official guidance, such as from the Information Commissioner's Office, the need to properly manage and process all personal data for reuse is addressed in codes of practice [12,13]. However, until recently, the vast majority of routine health data made available for secondary use was from coded records. As such, clinical free-text data are among the emerging research datasets, such as genomic and imaging data, and warrant further attention to develop official guidance to simultaneously safeguard individual privacy and maximize data utility for research in the public interest.

Our literature review revealed a variety of published studies that used data derived from free-text sources (eg, clinical notes and referral letters), sometimes in conjunction with coded health records and other datasets. The majority of these were conducted through established systems (eg, THIN and CRIS) specializing in making clinical free-text data available for approved research, in accordance with their operating and governance models. There were also some studies not associated with one of these systems. In many cases, the authors properly reported on how data were accessed and provided information on regulatory and governance approvals in place. However, there was also inconsistency and a lack of information in some studies, although this might reflect reporting rather than practice [79]. It is not always easy to ensure that authors provide governance details, but at least in studies from the established systems, this is a matter for publication policies that are part of their good governance practice. However, it should also be addressed by journal editors to ensure that there is a section providing transparency on methodology and data governance where free-text data (or person-based data) have been used.

There are a variety of automated methods to deidentify or extract data from free-text clinical documents for reuse in research or other purposes beyond direct clinical care. Our review showed that blacklisting methods are used by THIN and CRIS (and previously used by CPRD) and whitelisting methods are used by SAIL. THIN and SAIL operate in the higher education sector, whereas CRIS operates behind an NHS Trust firewall. All 3 institutes operate a data safe haven model, rather than external data release, and all have data governance models in place. From engagement with clinical text-mining researchers, it is clear that although the state-of-the-art systems achieve performance comparable with manual deidentification, there are still uncertainties about the efficacy of blacklisting methods and acceptable thresholds of reliability. Further work is needed to address this issue. However, whitelisting is not a panacea, as it also relies on algorithms that must be tested on identifiable data and can risk over- or underextracting data.

We have noted that using only small samples of identifiable free-text data sources can lead to bias. The creation of a free-text databank donated on a voluntary basis for developmental work would help to advance this field of work. At the same time, there is a need to minimize the need for human access to clinical free-text data through continual improvement in technological NLP methods for deidentification of clinical narratives. Further developments are also needed in other areas of text processing so that free-text data can be more efficiently converted to a structured format before reuse. Although this is advantageous in terms of risk minimization, it risks the loss of information richness that could be needed for qualitative research, exploring areas such as language, culture, and patient stories in the full free-text data, and reduces the future possible uses of the text data, with new extractions needed for subsequent research studies.

The clear need for patient involvement and engagement in the development of reusing free-text data was highlighted in the events with researchers and the public. This included a call for public-facing activities to be ongoing, inclusive, and transparent. Public delegates at the Turing Institute event and the consumer panel were generally positive about using free-text data for research, and this is in line with previous work with the Brighton Citizen's Jury (in 2018). This jury suggested that people's concerns could be mitigated with comprehensive patient-facing information about how, when, and under what conditions patients' free-text data might be used for research. Furthermore, patients should be involved as key stakeholders throughout all stages of the research process. The public also expects that researchers are committed to a culture of continuous improvement of methods for coding, anonymizing, processing, and safeguarding clinical free-text data [15].

Participant consent was a theme discussed at the clinical text-mining researcher workshop and again at the Turing Institute event, as it was at the Brighton Citizens' Jury. There was a recognition that it would not generally be practicable to seek opt-in consent to incorporate deidentified free-text extracts into data safe havens for reuse in research. The possibilities of opt-out consent were also explored and were considered favorable if they could be enacted in accordance with existing patient data guidelines. In England, there is a national opt-out where patients can choose not to have their data used for purposes beyond their direct care. However, in practice, this only applies to situations where the reuse of the data relies on s251 approval to set aside the CLDC [80]. As such, it is not a comprehensive opt-out mechanism and is not equipped to allow someone to specify a particular data format. Working in Wales, SAIL Databank operates its own opt-out mechanism by allowing individuals to inform their GP that they do not wish their data to be provided to the databank; again, this is an all-or-nothing option. More granular optouts would present difficult challenges in defining data content for inclusion and exclusion. Nevertheless, the mechanisms in place do provide individuals with choices they can exercise in relation to their patient records.

The panel discussion at the Turing Institute event stressed the importance of operating within existing frameworks for all data, not singling out clinical free-text data as fundamentally different. Although challenges remain for meaningful opt-out consent, mitigation can be provided via clear privacy notices, lawful bases for data processing, and transparent information on safeguards and their limitations. However, although the group discussions highlighted the great importance of being able to make good use of clinical free-text data, we are still left with the situation that the deidentification (blacklisting) of free-text data is imperfect, and extracting only certain elements (whitelisting) produces only a study-specific data extract. Although established data systems making free-text extracts available in one form or another have data governance controls in place, there are cases where further measures are needed at the project level. Two examples are studies that are not associated with established systems such that the data might be subject to external release (ie, to be held outside a data management infrastructure) and studies that need to use free-text data for qualitative analysis, where having the data translated into coded form would not be adequate.

Recommendations

In addition to good practice in handling person-based data in general, we propose the following set of recommendations and suggestions for further work to operate in accordance with, and augment, existing research and information governance frameworks:

- There is a need for clear regulatory guidance on data governance for the reuse of clinical free-text data, taking into account factors including whether the data are deidentified through blacklisting or extracted via whitelisting; the native data are housed at their original source or safe havens or exported to external users, and the extracts are in a coded or raw free-text form.
- Patients and the public should be better informed about free-text data flows and uses, including the availability of opt-out arrangements for the reuse of patient data for research in accordance with jurisdictional policies. This requires commitment from researchers and the public.
- The use of free-text data and results of studies should be publicized to all stakeholders, but particularly patients, so that the public can see that free-text data can bring additional benefits. This could take the form of regularly updated case studies in a central location, including information about the data used and the findings of the study.
- Further improvements are needed in deidentification and extraction algorithms with a better understanding of the relationships between the accuracy of deidentification and thresholds of reidentification risk. This could be facilitated by a national challenge with public demonstrations on how free-text data are deidentified.
- Owing to the current uncertainties, blacklisted clinical free-text data should be treated as potentially identifiable and access for research restricted to accredited data safe havens, unless a thorough review is conducted before release.
- Organizations providing access to clinical free-text extracts should stipulate in their publication policies a data governance statement to be placed in publications, and journals should require a suitable statement for all studies using data derived from free text.
- The specifics around the creation of a databank of donated clinical free-text data to support the construction of deidentification and extraction of algorithms should be explored; this would necessarily include further public engagement and government and NHS commitment.

Limitations

We acknowledge that the TexGov study was subject to limitations. It was a small study of a few months in duration and, as such, could not be comprehensive in scope. It was limited to the United Kingdom, and the governance of data use may differ in other jurisdictions. However, we believe that the principles will be largely applicable beyond the United Kingdom. We were unable to engage specifically with data providers or clinicians who created the data to gauge their viewpoints. This could be a focus of future work.

Conclusions

It is well known that clinical free-text data represent a rich resource for reuse in research, but that there are particular challenges in working with unstructured data to safeguard privacy and maximize data utility. We have shown that lessons can be learned from established systems providing access to data derived from clinical free text and that the views of text-mining researchers and members of the public provide valuable insights. We present the new knowledge gained in this unique study in the form of a position paper to work toward the development of data governance standards for the reuse of free-text data. While recognizing that free-text data are not fundamentally different from other patient data and the need to work within existing data governance frameworks, we propose that there is a need to develop the TexGov recommendations, with commitment and investment, to expand and assure the safe use of free-text data for public benefit.

Acknowledgments

TexGov was funded by the Engineering and Physical Sciences Research Council via Healtex, the UK health care text analytics research network (grant number EP/N027280/1).

Conflicts of Interest

None declared.

References

1. Data Linkage Western Australia. 2019. Enabling Health & Medical Research in Western Australia URL: https://www.datalinkage-wa.org.au/ [accessed 2019-10-18]
2. SAIL Databank - The Secure Anonymised Information Linkage. 2019. Secure Anonymised Information Linkage Databank URL: https://saildatabank.com/saildata/sail-datasets/ [accessed 2019-10-18]
3. Institute for Clinical Evaluative Sciences. 2019. About ICES URL: https://www.ices.on.ca/About-ICES [accessed 2019-10-18]
4. University of Manitoba. 2019. About the Manitoba Centre for Health Policy (MCHP) URL: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/about.html [accessed 2019-10-18]
5. General Data Protection Regulation (GDPR). 2019. URL: https://gdpr-info.eu/ [accessed 2019-10-18]
6. Kushida C, Nichols D, Jadrnicek R, Miller R, Walsh J, Griffin K. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. Med Care 2012 Jul;50(Suppl):S82-101 [FREE Full text] [doi: 10.1097/MLR.0b013e3182585355] [Medline: 22692265]
7. Meystre S, Friedlin F, South B, Shen S, Samore M. Automatic de-identification of textual documents in the electronic health record: a review of recent research. BMC Med Res Methodol 2010 Aug 2;10:70 [FREE Full text] [doi: 10.1186/1471-2288-10-70] [Medline: 20678228]
8. Fonferko-Shadrach B, Lacey AS, Roberts A, Akbari A, Thompson S, Ford DV, et al. Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the ExECT (extraction of epilepsy clinical text) system. BMJ Open 2019;9(4):e023232 [FREE Full text] [doi: 10.1136/bmjopen-2018-023232] [Medline: 30940752]
9. UK Legislation. 2018. Data Protection Act 2018 URL: http://www.legislation.gov.uk/ukpga/2018/12/notes/division/6/index.htm [accessed 2019-10-18]
10. UK Legislation. 1998. Human Rights Act 1998 URL: http://www.legislation.gov.uk/ukpga/1998/42/contents [accessed 2019-10-18]
11. UK Statistics. 2017. A Manual for Caldicott Guardians: Produced by the UK Caldicott Guardian Council URL: http://static.ukcgc.uk/docs/AManualForCaldicottGuardians.pdf [accessed 2019-10-18]
12. Information Commissioner's Office. 2012. Anonymisation: Managing Data Protection Risk Code of Practice URL: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf [accessed 2019-10-18]
13. Information Commissioner's Office. 2011. Data Sharing Code of Practice URL: https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf [accessed 2019-10-18]
14. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
15. Ford E, Oswald M, Hassan L, Bozentko K, Nenadic G, Cassell J. Should free-text data in electronic medical records be shared for research? A citizens' jury study in the UK. J Med Ethics 2020 May 26;46(6):367-377 [FREE Full text] [doi: 10.1136/medethics-2019-105472]
16. Government of UK. 2019. National Data Guardian URL: https://www.gov.uk/government/organisations/national-data-guardian [accessed 2019-10-18]
17. Understanding Patient Data. 2019. URL: https://understandingpatientdata.org.uk/ [accessed 2019-10-18]
18. Use My Data. 2019. URL: http://www.usemydata.org/?LMCL=kGltKL [accessed 2019-10-18]
19. NHS England. 2019. URL: https://www.england.nhs.uk/ [accessed 2019-10-18]
20. Jones K, McNerney CL, Ford DV. Involving consumers in the work of a data linkage research unit. Int J Consum Stud 2013 Oct 7;38(1):45-51 [FREE Full text] [doi: 10.1111/ijcs.12062] [Medline: 24843434]
21. Bourke A, Dattani HM, Robinson M. Feasibility study and methodology to create a quality-evaluated database of primary care data. Inform Prim Care 2004;12(3):171-177 [FREE Full text] [doi: 10.14236/jhi.v12i3.124] [Medline: 15606990]

22. Blak B, Thompson M, Dattani H, Bourke A. Generalisability of the health improvement network (THIN) database: demographics, chronic disease prevalence and mortality rates. Inform Prim Care 2011;19(4):251-255 [FREE Full text] [doi: 10.14236/jhi.v19i4.820] [Medline: 22828580]
23. Hall GC. Validation of death and suicide recording on the THIN UK primary care database. Pharmacoepidemiol Drug Saf 2009 Mar;18(2):120-131. [doi: 10.1002/pds.1686] [Medline: 19109801]
24. Ruigómez A, Martín-Merino E, Rodríguez LA. Validation of ischemic cerebrovascular diagnoses in the health improvement network (THIN). Pharmacoepidemiol Drug Saf 2010 Jun;19(6):579-585. [doi: 10.1002/pds.1919] [Medline: 20131328]
25. Martín-Merino E, Fortuny J, Rivero E, García-Rodríguez LA. Validation of diabetic retinopathy and maculopathy diagnoses recorded in a UK primary care database. Diabetes Care 2012 Apr;35(4):762-767 [FREE Full text] [doi: 10.2337/dc11-2069] [Medline: 22357184]
26. Woodman J, Allister J, Rafi I, de Lusignan S, Belsey J, Petersen I, RCGP Multisite Safeguarding Audit. A simple approach to improve recording of concerns about child maltreatment in primary care records: developing a quality improvement intervention. Br J Gen Pract 2012 Jul;62(600):e478-e486 [FREE Full text] [doi: 10.3399/bjgp12X652346] [Medline: 22781996]
27. Soriano LC, Soriano-Gabarró M, Rodríguez LA. Validity and completeness of colorectal cancer diagnoses in a primary care database in the United Kingdom. Pharmacoepidemiol Drug Saf 2016 Apr;25(4):385-391 [FREE Full text] [doi: 10.1002/pds.3877] [Medline: 26436320]
28. Soriano LC, López-Garcia E, Schulze-Rath R, Rodríguez LA. Incidence, treatment and recurrence of endometriosis in a UK-based population analysis using data from the health improvement network and the hospital episode statistics database. Eur J Contracept Reprod Health Care 2017 Oct;22(5):334-343. [doi: 10.1080/13625187.2017.1374362] [Medline: 28954555]
29. Cea-Soriano L, García-Rodríguez LA, Brodovicz KG, Gonzalez EM, Bartels DB, Hernández-Díaz S. Safety of non-insulin glucose-lowering drugs in pregnant women with pre-gestational diabetes: a cohort study. Diabetes Obes Metab 2018 Jul;20(7):1642-1651. [doi: 10.1111/dom.13275] [Medline: 29498473]
30. Stewart R, Soremekun M, Perera G, Broadbent M, Callard F, Denis M, et al. The south London and Maudsley NHS foundation trust biomedical research centre (SLAM BRC) case register: development and descriptive data. BMC Psychiatry 2009 Aug 12;9(1):51 [FREE Full text] [doi: 10.1186/1471-244X-9-51] [Medline: 19674459]
31. Perera G, Broadbent M, Callard F, Chang C, Downs J, Dutta R, et al. Cohort profile of the South London and Maudsley NHS foundation trust biomedical research centre (SLaM BRC) case register: current status and recent enhancement of an electronic mental health record-derived data resource. BMJ Open 2016 Mar 1;6(3):e008721 [FREE Full text] [doi: 10.1136/bmjopen-2015-008721] [Medline: 26932138]
32. Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang C, Jackson RG, et al. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC Med Inform Decis Mak 2013 Jul 11;13(1):71 [FREE Full text] [doi: 10.1186/1472-6947-13-71] [Medline: 23842533]
33. Wu C, Chang C, Robson D, Jackson R, Chen S, Hayes RD, et al. Evaluation of smoking status identification using electronic health records and open-text information in a large mental health case register. PLoS One 2013;8(9):e74262 [FREE Full text] [doi: 10.1371/journal.pone.0074262] [Medline: 24069288]
34. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of adverse drug events from free text electronic patient records and information in a large mental health case register. PLoS One 2015;10(8):e0134208 [FREE Full text] [doi: 10.1371/journal.pone.0134208] [Medline: 26273830]
35. Kadra G, Stewart R, Shetty H, Jackson RG, Greenwood MA, Roberts A, et al. Extracting antipsychotic polypharmacy data from electronic health records: developing and evaluating a novel process. BMC Psychiatry 2015 Jul 22;15:166 [FREE Full text] [doi: 10.1186/s12888-015-0557-z] [Medline: 26198696]
36. Patel R, Jayatilleke N, Broadbent M, Chang CK, Foskett N, Gorrell G, et al. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. BMJ Open 2015 Sep 7;5(9):e007619 [FREE Full text] [doi: 10.1136/bmjopen-2015-007619] [Medline: 26346872]
37. Patel R, Lloyd T, Jackson R, Ball M, Shetty H, Broadbent M, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. BMJ Open 2015 May 21;5(5):e007504 [FREE Full text] [doi: 10.1136/bmjopen-2014-007504] [Medline: 25998036]
38. Patel R, Wilson R, Jackson R, Ball M, Shetty H, Broadbent M, et al. Association of cannabis use with hospital admission and antipsychotic treatment failure in first episode psychosis: an observational study. BMJ Open 2016 Mar 3;6(3):e009888 [FREE Full text] [doi: 10.1136/bmjopen-2015-009888] [Medline: 26940105]
39. Taylor CL, Stewart R, Ogden J, Broadbent M, Pasupathy D, Howard LM. The characteristics and health needs of pregnant women with schizophrenia compared with bipolar disorder and affective psychoses. BMC Psychiatry 2015 Apr 17;15(1):88 [FREE Full text] [doi: 10.1186/s12888-015-0451-8] [Medline: 25886140]
40. Knapp M, Chua K, Broadbent M, Chang CK, Fernandez JL, Milea D, et al. Predictors of care home and hospital admissions and their costs for older people with Alzheimer's disease: findings from a large London case register. BMJ Open 2016 Nov 18;6(11):e013591 [FREE Full text] [doi: 10.1136/bmjopen-2016-013591] [Medline: 27864252]
41. Colling C, Evans L, Broadbent M, Chandran D, Craig TJ, Kolliakou A, et al. Identification of the delivery of cognitive behavioural therapy for psychosis (CBTp) using a cross-sectional sample from electronic health records and open-text

information in a large UK-based mental health case register. BMJ Open 2017 Jul 17;7(7):e015297 [FREE Full text] [doi: 10.1136/bmjopen-2016-015297] [Medline: 28716789]

42. Das-Munshi J, Chang C, Dutta R, Morgan C, Nazroo J, Stewart R, et al. Ethnicity and excess mortality in severe mental illness: a cohort study. Lancet Psychiatry 2017 May;4(5):389-399 [FREE Full text] [doi: 10.1016/S2215-0366(17)30097-4] [Medline: 28330589]
43. Iqbal E, Mallah R, Rhodes D, Wu H, Romero A, Chang N, et al. ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. PLoS One 2017;12(11):e0187121 [FREE Full text] [doi: 10.1371/journal.pone.0187121] [Medline: 29121053]
44. Jackson R, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the clinical record interactive search comprehensive data extraction (CRIS-CODE) project. BMJ Open 2017 Jan 17;7(1):e012012 [FREE Full text] [doi: 10.1136/bmjopen-2016-012012] [Medline: 28096249]
45. Das-Munshi J, Chang C, Schofield P, Stewart R, Prince MJ. Depression and cause-specific mortality in an ethnically diverse cohort from the UK: 8-year prospective study. Psychol Med 2019 Jul;49(10):1639-1651 [FREE Full text] [doi: 10.1017/S0033291718002210] [Medline: 30180917]
46. Downs J, Dean H, Lechler S, Sears N, Patel R, Shetty H, et al. Negative symptoms in early-onset psychosis and their association with antipsychotic treatment failure. Schizophr Bull 2019 Jan 1;45(1):69-79 [FREE Full text] [doi: 10.1093/schbul/sbx197] [Medline: 29370404]
47. Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. Sci Rep 2018 May 9;8(1):7426 [FREE Full text] [doi: 10.1038/s41598-018-25773-2] [Medline: 29743531]
48. Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: a general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. J Am Med Inform Assoc 2018 May 1;25(5):530-537 [FREE Full text] [doi: 10.1093/jamia/ocx160] [Medline: 29361077]
49. Jackson R, Kartoglu I, Stringer C, Gorrell G, Roberts A, Song X, et al. CogStack - experiences of deploying integrated information retrieval and extraction services in a large national health service foundation trust hospital. BMC Med Inform Decis Mak 2018 Jun 25;18(1):47 [FREE Full text] [doi: 10.1186/s12911-018-0623-9] [Medline: 29941004]
50. Clinical Practice Research Datalink. 2019. Medicines and Healthcare Products Regulatory Agency URL: https://www.cprd.com/ [accessed 2019-10-18]
51. Wurst K, Ephross S, Loehr J, Clark DW, Guess HA. The utility of the general practice research database to examine selected congenital heart defects: a validation study. Pharmacoepidemiol Drug Saf 2007 Aug;16(8):867-877. [doi: 10.1002/pds.1431] [Medline: 17563909]
52. Charlton R, Weil J, Cunnington MC, de Vries CS. Identifying major congenital malformations in the UK general practice research database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. Drug Saf 2010 Sep 1;33(9):741-750. [doi: 10.2165/11536820-000000000-00000] [Medline: 20701407]
53. Devine S, West S, Andrews E, Tennis P, Hammad TA, Eaton S, et al. The identification of pregnancies within the general practice research database. Pharmacoepidemiol Drug Saf 2010 Jan;19(1):45-50. [doi: 10.1002/pds.1862] [Medline: 19823973]
54. Tate A, Martin A, Ali A, Cassell JA. Using free text information to explore how and when GPs code a diagnosis of ovarian cancer: an observational study using primary care records of patients with ovarian cancer. BMJ Open 2011 Mar 23;1(1):e000025 [FREE Full text] [doi: 10.1136/bmjopen-2010-000025] [Medline: 22021731]
55. Shah A, Martinez C, Hemingway H. The freetext matching algorithm: a computer program to extract diagnoses and causes of death from unstructured text in electronic health records. BMC Med Inform Decis Mak 2012 Aug 7;12:88 [FREE Full text] [doi: 10.1186/1472-6947-12-88] [Medline: 22870911]
56. Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. PLoS One 2012;7(1):e30412 [FREE Full text] [doi: 10.1371/journal.pone.0030412] [Medline: 22276193]
57. Ford E, Nicholson A, Koeling R, Tate A, Carroll J, Axelrod L, et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? BMC Med Res Methodol 2013 Aug 21;13:105 [FREE Full text] [doi: 10.1186/1471-2288-13-105] [Medline: 23964710]
58. Nicholson A, Ford E, Davies KA, Smith HE, Rait G, Tate AR, et al. Optimising use of electronic health records to describe the presentation of rheumatoid arthritis in primary care: a strategy for developing code lists. PLoS One 2013;8(2):e54878 [FREE Full text] [doi: 10.1371/journal.pone.0054878] [Medline: 23451024]
59. Ford E, Carroll J, Smith H, Davies K, Koeling R, Petersen I, et al. What evidence is there for a delay in diagnostic coding of RA in UK general practice records? An observational study of free text. BMJ Open 2016 Jun 28;6(6):e010393 [FREE Full text] [doi: 10.1136/bmjopen-2015-010393] [Medline: 27354069]
60. Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. Is omission of free text records a possible source of data loss and bias in clinical practice research datalink studies? A case-control study. BMJ Open 2016 May 13;6(5):e011664 [FREE Full text] [doi: 10.1136/bmjopen-2016-011664] [Medline: 27178981]

61. Karystianis G, Sheppard T, Dixon WG, Nenadic G. Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database. BMC Med Inform Decis Mak 2016 Mar 9;16(1):18 [FREE Full text] [doi: 10.1186/s12911-016-0255-x] [Medline: 26860263]
62. Maguire A, Johnson M, Denning DW, Ferreira GL, Cassidy A. Identifying rare diseases using electronic medical records: the example of allergic bronchopulmonary aspergillosis. Pharmacoepidemiol Drug Saf 2017 Jul;26(7):785-791 [FREE Full text] [doi: 10.1002/pds.4204] [Medline: 28370596]
63. Zhou S, Rahman M, Atkinson M, Brophy S. Mining Textual Data From Primary Healthcare Records: Automatic Identification of Patient Phenotype Cohorts. In: Proceedings of the International Joint Conference on Neural Networks. 2014 Presented at: IJCNN'14; July 6-11, 2014; Beijing, China p. 3621-3627. [doi: 10.1109/ijcnn.2014.6889494]
64. Premaratne U, Marks G, Austin E, Burney P. A reliable method to retrieve accident & emergency data stored on a free-text basis. Respir Med 1997 Mar;91(2):61-66 [FREE Full text] [doi: 10.1016/s0954-6111(97)90069-x] [Medline: 9122513]
65. Brown PJ, Sönksen P. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. J Am Med Inform Assoc 2000;7(4):392-403 [FREE Full text] [doi: 10.1136/jamia.2000.0070392] [Medline: 10887167]
66. Young J, Goldfrad C, Rowan K. Development and testing of a hierarchical method to code the reason for admission to intensive care units: the ICNARC coding method. Intensive care national audit & research centre. Br J Anaesth 2001 Oct;87(4):543-548 [FREE Full text] [doi: 10.1093/bja/87.4.543] [Medline: 11878722]
67. Anandarajah S, Tai T, de Lusignan S, Stevens P, O'Donoghue D, Walker M, et al. The validity of searching routinely collected general practice computer data to identify patients with chronic kidney disease (CKD): a manual review of 500 medical records. Nephrol Dial Transplant 2005 Oct;20(10):2089-2096. [doi: 10.1093/ndt/gfi006] [Medline: 16030033]
68. Flynn RW, Macdonald TM, Schembri N, Murray GD, Doney AS. Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes. Pharmacoepidemiol Drug Saf 2010 Aug;19(8):843-847. [doi: 10.1002/pds.1981] [Medline: 20602346]
69. Flynn RW, MacDonald TM, Murray GD, Ferguson C, Shah K, Doney AS. The tayside stroke cohort: exploiting advanced regional medical informatics to create a region-wide database for studying the pharmacoepidemiology of stroke. Pharmacoepidemiol Drug Saf 2010 Jul;19(7):737-744. [doi: 10.1002/pds.1958] [Medline: 20583205]
70. Anwar MN, Oakes MP. Data mining of audiology patient records: factors influencing the choice of hearing aid type. BMC Med Inform Decis Mak 2012 Apr 30;12(Suppl 1):S6 [FREE Full text] [doi: 10.1186/1472-6947-12-S1-S6] [Medline: 22595091]
71. Coleman J, McDowell SE, Ferner RE. Dose omissions in hospitalized patients in a UK hospital: an analysis of the relative contribution of adverse drug reactions. Drug Saf 2012 Aug 1;35(8):677-683. [doi: 10.1007/bf03261964] [Medline: 22734657]
72. Donaldson L, Panesar SS, Darzi A. Patient-safety-related hospital deaths in England: thematic analysis of incidents reported to a national database, 2010-2012. PLoS Med 2014 Jun;11(6):e1001667 [FREE Full text] [doi: 10.1371/journal.pmed.1001667] [Medline: 24959751]
73. Hayward RA, Chen Y, Croft P, Jordan KP. Presentation of respiratory symptoms prior to diagnosis in general practice: a case-control study examining free text and morbidity codes. BMJ Open 2015 Jun 12;5(6):e007355 [FREE Full text] [doi: 10.1136/bmjopen-2014-007355] [Medline: 26070795]
74. Willits I, Burn J, Cole H, Hoare T, Sims A. What proportions of focal liver lesions detected by unenhanced ultrasound are inconclusive? Ultrasound 2015 May;23(2):78-84 [FREE Full text] [doi: 10.1177/1742271X14562995] [Medline: 25949268]
75. López-García P, Kreuzthaler M, Schulz S, Scherr D, Daumke P, Markó K, et al. SEMCARE: multilingual semantic search in semi-structured clinical data. Stud Health Technol Inform 2016;223:93-99. [Medline: 27139390]
76. Sheaff R, Halliday J, Byng R, Øvretveit J, Exworthy M, Peckham S, et al. Bridging the discursive gap between lay and medical discourse in care coordination. Sociol Health Illn 2017 Sep;39(7):1019-1034. [doi: 10.1111/1467-9566.12553] [Medline: 28349619]
77. Jones KH, Ford DV, Jones C, Dsilva R, Thompson S, Brooks CJ, et al. A case study of the secure anonymous information linkage (SAIL) gateway: a privacy-protecting remote access system for health-related research and evaluation. J Biomed Inform 2014 Aug;50:196-204 [FREE Full text] [doi: 10.1016/j.jbi.2014.01.003] [Medline: 24440148]
78. Health Research Authority. 2019. Research Passport URL: https://www.hra.nhs.uk/planning-and-improving-research/best-practice/research-passport/ [accessed 2019-10-01]
79. Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, et al. Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances. J Biomed Inform 2018 Dec;88:11-19 [FREE Full text] [doi: 10.1016/j.jbi.2018.10.005] [Medline: 30368002]
80. NHS Digital. 2019. National Data Opt-Out URL: https://digital.nhs.uk/services/national-data-opt-out [accessed 2019-10-05]

Abbreviations

CLDC: common law duty of confidentiality
CPRD: Clinical Practice Research Datalink
CRIS: Clinical Records Interactive Search

DPA: Data Protection Act
EHR: electronic health record
GDPR: General Data Protection Regulation
GP: general practitioner
IGRP: Information Governance Review Panel
NHS: National Health Service
NLP: natural language processing
PID: person-identifiable data
SAIL: Secure Anonymised Information Linkage
SLaM: South London and Maudsley NHS Foundation Trust
THIN: the Health Improvement Network
UCL: University College London
WoS: Web of Science

Edited by G Eysenbach; submitted 22.10.19; peer-reviewed by H Daniels, A Ekeland; comments to author 27.01.20; revised version received 06.03.20; accepted 23.03.20; published 29.06.20.

Please cite as:
Jones KH, Ford EM, Lea N, Griffiths LJ, Hassan L, Heys S, Squires E, Nenadic G
Toward the Development of Data Governance Standards for Using Clinical Free-Text Data in Health Research: Position Paper
J Med Internet Res 2020;22(6):e16760
URL: http://www.jmir.org/2020/6/e16760/
doi:10.2196/16760
PMID:32597785

Original Paper

Healthcare Research and Analytics Data Infrastructure Solution: A Data Warehouse for Health Services Research

Bunyamin Ozaydin[1], MSEE, PhD; Ferhat Zengul[1], MBA, MAC, PhD; Nurettin Oner[1], MSHA; Sue S Feldman[1], RN, MEd, PhD

University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Bunyamin Ozaydin, MSEE, PhD
University of Alabama at Birmingham
SHPB 590H
1720 2nd Ave S
Birmingham, AL, 35294-1212
United States
Phone: 1 205 996 7242
Email: bozaydin@uab.edu

Abstract

Background: Health services researchers spend a substantial amount of time performing integration, cleansing, interpretation, and aggregation of raw data from multiple public or private data sources. Often, each researcher (or someone in their team) duplicates this effort for their own project, facing the same challenges and experiencing the same pitfalls discovered by those before them.

Objective: This paper described a design process for creating a data warehouse that includes the most frequently used databases in health services research.

Methods: The design is based on a conceptual iterative process model framework that utilizes the sociotechnical systems theory approach and includes the capacity for subsequent updates of the existing data sources and the addition of new ones. We introduce the theory and the framework and then explain how they are used to inform the methodology of this study.

Results: The application of the iterative process model to the design research process of problem identification and solution design for the Healthcare Research and Analytics Data Infrastructure Solution (HRADIS) is described. Each phase of the iterative model produced end products to inform the implementation of HRADIS. The analysis phase produced the problem statement and requirements documents. The projection phase produced a list of tasks and goals for the *ideal* system. Finally, the synthesis phase provided the process for a plan to implement HRADIS. HRADIS structures and integrates data dictionaries provided by the data sources, allowing the creation of dimensions and measures for a multidimensional business intelligence system. We discuss how HRADIS is complemented with a set of data mining, analytics, and visualization tools to enable researchers to more efficiently apply multiple methods to a given research project. HRADIS also includes a built-in security and account management framework for data governance purposes to ensure customized authorization depending on user roles and parts of the data the roles are authorized to access.

Conclusions: To address existing inefficiencies during the obtaining, extracting, preprocessing, cleansing, and filtering stages of data processing in health services research, we envision HRADIS as a full-service data warehouse integrating frequently used data sources, processes, and methods along with a variety of data analytics and visualization tools. This paper presents the application of the iterative process model to build such a solution. It also includes a discussion on several prominent issues, lessons learned, reflections and recommendations, and future considerations, as this model was applied.

(J Med Internet Res 2020;22(6):e18579) doi:10.2196/18579

KEYWORDS

health services research; data warehousing; iterative process model; systems analysis and design; data integration

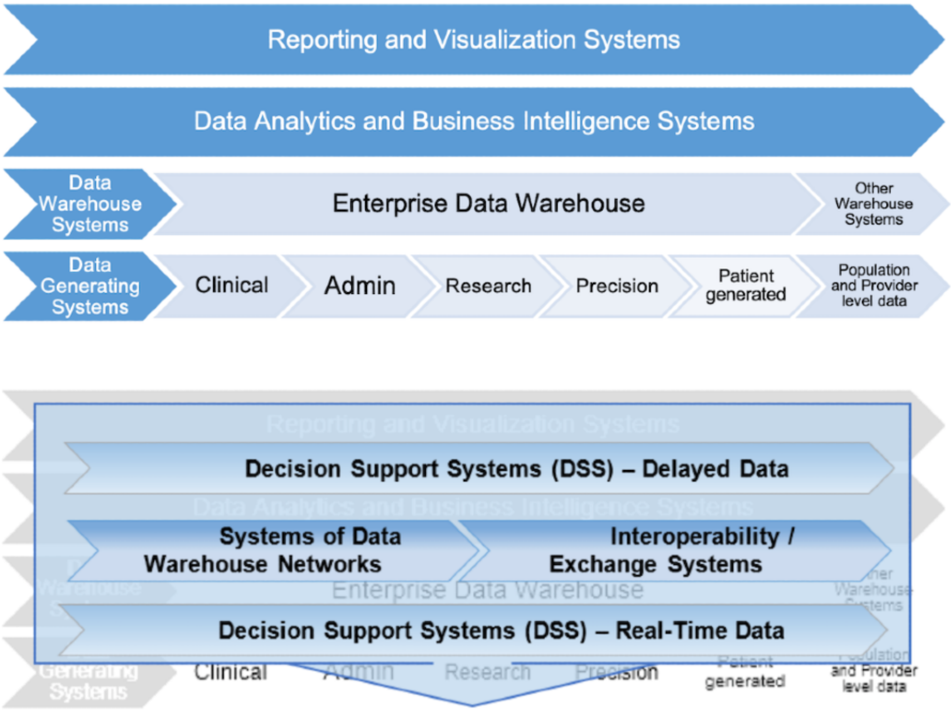
Introduction

There are a variety of data sources most frequently used for health services research, a multidisciplinary research field that investigates the implications of factors such as social determinants, organizational structures and processes, technologies, financing and reimbursement, individual choices and behaviors on the access and quality of health care delivery, and overall health and well-being of individuals [1]. Most of the data sources for health services research are provided by the Centers for Medicare and Medicaid Services (CMS); however, there are also data sources provided by other government agencies and nonprofit or for-profit data providers. Health services researchers, especially those using secondary data, can expand their research analytics by using merged datasets for health services research. In the absence of a single data warehouse from which to retrieve and analyze data from previously disparate datasets, health services researchers are forced to perform separate and often redundant data-related tasks on each individual dataset. Anecdotal reports suggest that researchers spend as much as 60% of their time on data preparation. At best, we can describe the current data-related processes as inefficient, costly, time-consuming, and cumbersome [2]. Moreover, the current uncoordinated and isolated efforts on these disparate datasets can be wasteful as they may generate research findings that are not reproducible or sometimes misleading because of the unaddressed inherent problems within these datasets. Furthermore, without the needed information technology (IT) infrastructure, analytics, and data visualization tools, the potential of the ever-growing health-related big data accumulated in these disparate datasets would still be untapped [3]. Therefore, there is a need for a cyberinfrastructure that integrates these disparate databases in a secure and consistent manner and provides the necessary analytics and visualization tools.

Background: Systems Around Health Data

In its life cycle, health-related data mainly move through four types of systems, as indicated in the top part of Figure 1 (adapted from the study by Ozaydin et al [4]). Patient-level data are usually generated in one of the operational systems that fall into categories of clinical, administrative, research, and precision medicine systems and systems that manage medical devices that patients use. The clinical systems include everything that is part of the electronic health record (EHR) and systems dealing with laboratories, imaging, physician notes, medications, histories, procedures, and diagnoses, regardless of whether or not they are part of the EHR. Administrative systems include admittance-discharge-transfer; billing, scheduling, and claims systems; as well as systems that are not specific to health care, such as systems that manage human resources and payroll. The research-related health data are generated by the systems for clinical research, clinical trials, and various registries. Furthermore, there are systems generating precision medicine data, such as genomics, phonemics, and microbiome, and systems where patient-generated data are generated, such as mobile health and telehealth systems, internet-of-things, and other data-generating medical devices, social media, and patient portals. After being created in one of the data-generating systems, the patient-level data are usually aggregated at an institutional enterprise data warehouse system. These data warehouses usually serve as the infrastructure on which institutional data analytics and business intelligence (BI) systems—based on which reporting and visualization systems, such as dashboards—run [4]. There are also other data warehouse systems outside of individual institutions, such as systems used for public health purposes [5,6].

Figure 1. Systems that deal with health-related data: First layer (top), second layer (bottom).



As indicated in the bottom part of Figure 1, there are also second-layer systems that operate between the systems shown in the top part of Figure 1. The second-layer systems include decision support systems, systems that provide interoperability between the first-layer systems (ie, interface engines), health information exchanges, and networks of institutional data warehouses. These second-layer systems are closer to data-generating systems if they are required to use real-time data. In the absence of the use of real-time data, these systems rely on delayed data provided by the data warehouse systems. As decision support systems mature, there are increasing expectations to provide their results back to the data-generating systems as close to real time as possible. To date, several data warehouse networks, such as the networks of informatics for integrating biology and the bedside (i2b2) systems called the shared health research information network, have been developed as data warehouse networks to integrate clinical and administrative data extracted from various systems of health care entities.

Need for the Healthcare Research and Analytics Data Infrastructure Solution

The systems described so far are mostly geared toward integrating patient-level electronic health, billing, and other administrative data to be used for clinical and translational research [2], without much focus on the organizational-level data. The first-layer systems also include systems that generate population- and provider-level data and data warehouse systems for them as indicated on the right-hand side of Figure 1. The population-level warehouse systems focus on epidemiological systems, systems managing national and regional indexes and surveys, and systems managing the Centers for Disease Control and Prevention databases. The provider-level warehouse systems focus on systems that manage data for health services administration, such as quality measures, satisfaction scores, inspections, financial performance, and services offered.

In addition to data warehouse networks for patient-level data, there have also been attempts to create integrated data repositories to include certain portions of the selected data sources for health services research for various purposes (ie, Research Data Assistance Center [ResDAC] [7] and Wharton research data services [WRDS] [8]). However, we could not find evidence of any mature platform that integrates all of the targeted data sources mentioned in Textbox 1 or of any effort to create such a platform in the literature. A majority of health services data continue to aggregate and evolve as isolated silos within various governmental or nongovernmental entities [2], with research efforts to interpret the data also operating in silos. In an era where the generation of data surpasses the efforts to extract meaning out of it, these uncoordinated silos of research efforts delay the necessary improvements in much-needed research efficiency. Therefore, enhancing health services research efficiency necessitates a platform that has the potential to integrate the disparate silos of datasets and research efforts.

Textbox 1. Data sources included in the first phase of the Healthcare Research and Analytics Data Infrastructure Solution.

- Centers for Medicare and Medicaid Services (CMS) Medicare cost reports
- CMS impact and final rule files
- Datasets from CMS Hospital Compare, including Hospital Consumer Assessment of Healthcare Providers and Systems
- Area health resources files
- American Hospital Association (AHA) annual survey
- AHA health information technology supplement
- Dartmouth Atlas
- Bureau of Labor Statistics

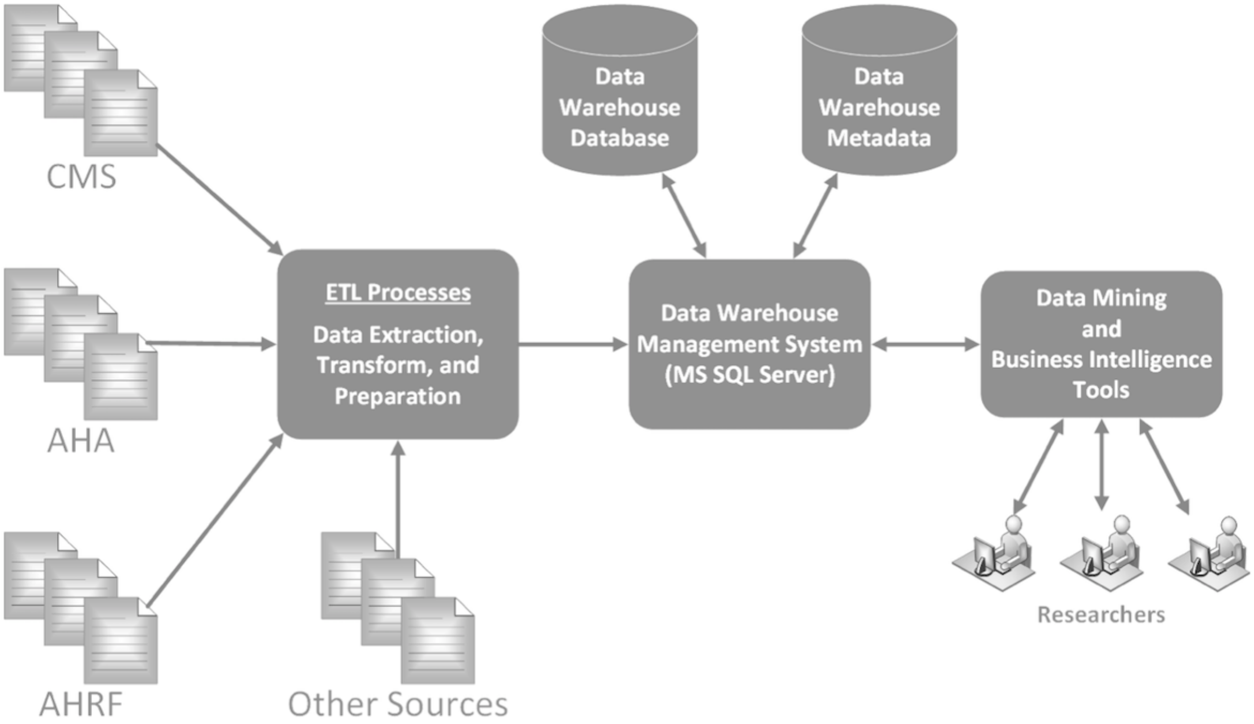
Methods

This section introduces data sources and architecture for the Healthcare Research and Analytics Data Infrastructure Solution (HRADIS) platform and the theoretical orientation and explains how the theoretical background is used to inform the methodology for this study.

Healthcare Research and Analytics Data Infrastructure Solution Data Sources and Architecture

To address the aforementioned need, this project aimed to generate a cyberinfrastructure by initially creating a data warehouse using the Microsoft SQL Server platform to integrate these frequently used health services data sources in a reliable, secure, and consistent manner and then to build a BI system that includes tools for data mining, analytics, and visualization, as depicted in Figure 2 (adapted from Kroenke and Auer [9]). The elements of HRADIS include data; metadata; procedures and applications of the data and metadata; other data tools; and users, groups, and data access policies. As shown in Figure 2, there are several different areas of data interaction. First, the data interact with the ETL (extract, transform, load) processes where data are prepared for storage in the data warehouse. Next, the data warehouse management system stores the data and metadata and handles data interaction between various other system tools and the stored data. Finally, the health services researchers interact with a graphical user interface to access the data through data mining and BI tools. The first phase of HRADIS hosts data from the data sources listed in Textbox 1.

Figure 2. The healthcare research and analytics data infrastructure solution architecture. CMS: Centers for Medicare and Medicaid Services; AHA: American Hospital Association; AHRF: area health resources files.



Some of the data sources listed in Textbox 1 have been made available through several research data centers such as ResDAC at the University of Minnesota, WRDS at the University of Pennsylvania, and the National Bureau of Economic Research (NBER) [10]. However, these research data centers do not include most of the data sources listed in Textbox 1. They primarily provide training and technical assistance on specific data sources such as CMS Medicare and Medicaid data in the case of ResDAC, focus on nonhealth care areas such as finance and business in the case of WRDS or improve the accessibility of existing data sources such as CMS Medicare cost reports (MCR) in the case of NBER. There are also commercial data centers, which provide reports on quality, finance, and inpatient and outpatient outcomes, for individual hospitals such as the American hospital directory [11] or hospital profiles such as Hospital-data [12]. However, these commercial data centers are not as comprehensive as, and some lack research focus when compared with HRADIS.

HRADIS is designed to be sustainable and scalable so that the inclusion of new data sources and updates of existing ones is efficient. This allows health services researchers to apply their models to updated data or data from new sources without having to merge new data to their research datasets.

A challenge exists in interpreting complex data dictionaries, layouts, and other metadata elements that accompany raw data to be able to identify and reliably extract parameters of interest for a given research project. To address this issue, HRADIS integrates metadata and ETL processes that utilize it to identify and extract parameters of interest based on how the parameters are defined by the metadata, rather than keeping metadata in a separate file repository or in an accompanying document warehouse [13]. To accomplish this, the source data are put through an initial phase of ETL tasks to populate database tables created based on the relational HRADIS data model in the entity relationship diagram format. Furthermore, HRADIS includes a second phase of ETL tasks to populate dimensions and measures that are created for the most frequently used parameters for more efficient performance of data mining, analytics, and visualization tasks based on the multidimensional HRADIS data model in a star schema diagram format. In other words, HRADIS benefits the advantages of both relational and dimensional models and their diagrammatical representations, as described in studies by Corral et al [14] and Schuff et al [15].

Currently, we have loaded MCR, AHA, Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), Dartmouth Atlas, Bureau of Labor Statistics (BLS), and parts of Hospital Compare datasets with HRADIS and have begun data extraction from multiple data sources for pilot projects. Although most of the data incorporated into the system come from publicly accessible sources, some of the data are restricted. As we populate HRADIS with data, we are also implementing the security and data governance components of the system. Security and data governance are important for users to access only the parts of the data for which they are authorized.

Sociotechnical Systems Theory

Although cyberinfrastructures can be designed and developed in a system-centric vacuum, the associated functionality must consider the role of the user and how the user will interact with the data housed in the cyberinfrastructure. A sociotechnical systems approach that takes into account the interaction between the human and the technology [16,17] is therefore appropriate because it promotes theoretical development while enabling system designers and developers to incorporate social awareness, organizational behavior, or other underrepresented domains, such as culture, which may be a critical component in system use. Many engineered system innovations fail in terms of adoption or use due to their lack of attention on human-technology interactions that are necessary and unavoidable [18,19]. An innovative cyberinfrastructure that introduces various changes to the existing practices would potentially fail if its psychosocial implications are not recognized. Moreover, in the current era, both the continuous coevolution of society and technology [20,21], especially the emergence of virtual organizations [22] that utilize telecommunication [23] or electronic-learning tools [24-26], and a surge in the amount of digital data (ie, big data) [27] create challenges for system developers in designing user-friendly, yet adaptive and sophisticated cyberinfrastructures. HRADIS considers the importance of intuitive user interfaces that are cognizant of the psychosocial and educational backgrounds of its users. It is worth noting that achieving the adaptability and sophistication with the simplicity that user-friendliness requires may necessitate more investment in the information systems design and development processes.

Iterative Process Model as a Conceptual Framework

Multiple databases are available, but remain disparate, making it difficult, if not impossible, for the health services researcher to conduct and collaborate on innovative and rigorous research that has currency and relevance. The literature provides evidence of the importance of the design theory in focusing on the design process in artifact development [28,29]. As such, we used a design science process framework to guide the design of an artifact that aims to improve data delivery to health services researchers such that practitioners more readily benefit from the insights and findings.

Design science is an essential component in information systems research that holds promise to improve research capabilities. Through artifact creation, health services researchers have immediate access to multiple and expanding datasets, offering opportunities for comparisons previously thought cumbersome and time-consuming. This artifact can be a construct, method, model, or instantiation [30]. This paper focused on the *method* used to design and develop a data warehouse for health services researchers.

Offermann et al [31] synthesized the design research process into 3 categories: (1) problem identification, (2) solution design, and (3) evaluation. This paper reports the information systems design *problem identification and solution design* categories only.

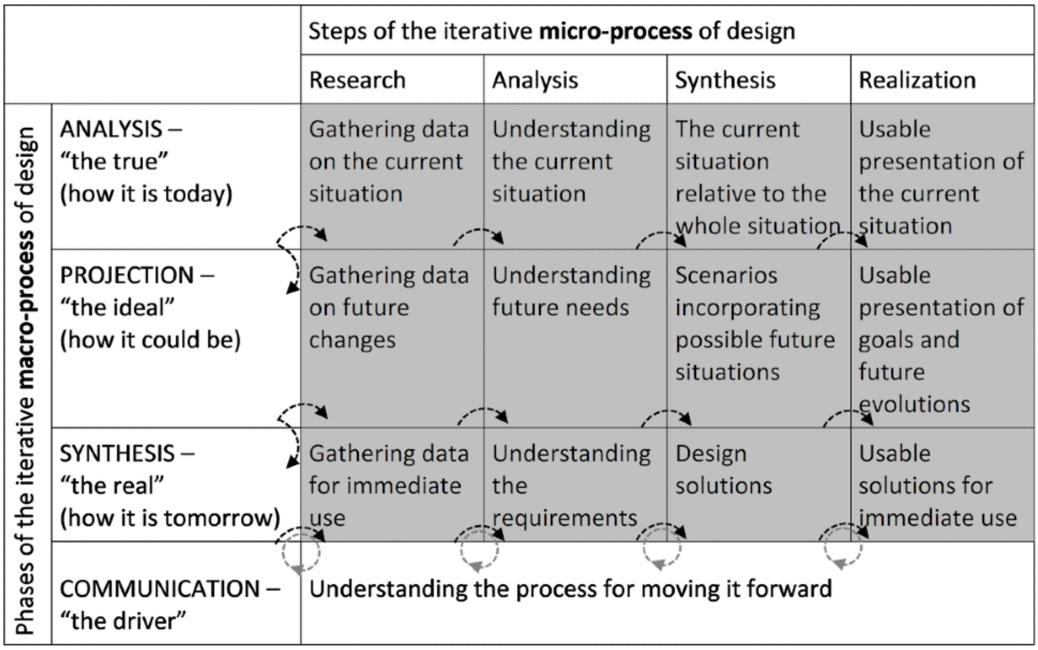
The literature offers multiple approaches for problem identification [32,33]. Historically, interviews were conducted with relevant end users so that designers could understand the issues as the users saw them. In addition, previous studies in the literature illuminate the problems that researchers look to solve. However, more recently, one needs to only read the headlines for problem identification: a lot of data, in many different places, accumulating very quickly. Some call this *big data*, but regardless of what label it is given, health services researchers are clamoring for efficient ways to cleanse, combine, analyze, and visualize the disparate datasets for ease of analysis, collaboration, action, and publication. Doing so holds promise to analyze and visualize combinations of data to reveal information that, when put into practice, can give their organization a competitive advantage.

Although there is much literature on design engineering information systems to accept data, there is very little literature on considerations to design solutions specific to disparate health datasets, or more simply stated, a *how to* approach [5,6].

Solution design is part of an evolutionary process that helps to operationalize solutions and general system analysis and design principles. The model proposed by Jonas [34], as shown in Table 1, is appropriate for use as a conceptual model in the design and development of a data warehouse for health services researchers because it allows for consideration of the entire process and encourages creative solution design. In Figure 3, 4 domains of design inquiry (ANALYSIS, PROJECTION, SYNTHESIS, and COMMUNICATION) are indicated as *phases* of the iterative macro process of design and are denoted with all capital letters. The 4 *steps* of the iterative microprocess of design (Research, Analysis, Synthesis, and Realization), on the other hand, are *denoted with first capital and the other lower-case letters*. Each previous microprocess *step* informs the next microprocess *step*, as indicated by the arrows. Similarly, the output from each macro process *phase* of design, which considers each step of the iterative microprocess of design, then informs the next macroprocess *phase*. The dotted lines on the arrows between the microprocess steps and the macroprocess phases denote that this is an iterative process. As COMMUNICATION is the driver for all macrophases and microsteps, this continual and iterative process is denoted by a dotted circular process arrow.

The 12 shaded sections in Figure 3 contain what occurs for each respective step. For example, in the ANALYSIS/Synthesis step, there needs to be an understanding of the current situation relative to the whole. The model is not prescriptive in exactly *how* this understanding occurs and thus allows for various types of individualized design processes. Within the first 2 phases of the iterative macroprocess (ANALYSIS and PROJECTION), the model allows for moving from Research (gathering data about the problems) to Analysis (understanding those problems) to Synthesis (expressing the problems from the perspective of looking at the system as a whole and assigning the problems into categories) and then Realization (presenting these problems as a problem/requirements statement). The difference between these first 2 phases is that during ANALYSIS, the model allows us to focus on the current problems and requirements, whereas during PROJECTION, the model allows for focusing on future problems and the requirements of an ideal system. In both cases, the end product is the presentation of the problems in their respective categories. In SYNTHESIS, all the knowledge learned from the previous 2 phases informs the Research, Analysis, Synthesis (or design), and Realization (or development and implementation) of the core functionalities of the entire system, that is, the first version of the system. We address how we interpret COMMUNICATION in the *Methodology* section.

Figure 3. The iterative process model.



		Steps of the iterative **micro-process** of design			
		Research	Analysis	Synthesis	Realization
Phases of the iterative **macro-process** of design	ANALYSIS – "the true" (how it is today)	Gathering data on the current situation	Understanding the current situation	The current situation relative to the whole situation	Usable presentation of the current situation
	PROJECTION – "the ideal" (how it could be)	Gathering data on future changes	Understanding future needs	Scenarios incorporating possible future situations	Usable presentation of goals and future evolutions
	SYNTHESIS – "the real" (how it is tomorrow)	Gathering data for immediate use	Understanding the requirements	Design solutions	Usable solutions for immediate use
	COMMUNICATION – "the driver"	Understanding the process for moving it forward			

Methodology

Consistent with the iterative process model, the first phase, ANALYSIS, is to understand the current situation with the data and then to realize a usable presentation of the current situation. This is accomplished through the following:

- Research: gathering disparate data, databases, their metadata, and the problems health services researchers currently face dealing with these data sources.
- Analysis: understanding the data, its structure, and metadata in each database and the domains of problems health services researchers face.
- Synthesis: merging these domains of problems and our understanding of the data and metadata from the perspective of the data warehouse project as a whole.
- Realization: creating an initial problem statement and requirements documentation for the project.

Table 1 presents a summary of this process relative to the ANALYSIS phase of the data warehouse project.

Once this first phase, ANALYSIS of how the data, metadata, disparate databases and their structures and evolution up to this point, and the problems around utilizing them for health services research, is appreciated, it is time to consider phase 2, PROJECTION, or what the ideal state of HRADIS would be. This second macrophase in the iterative design process is concerned with the future needs of the project. First, in the Research step, we gather data about the additional problems and requirements that may surface as a result of Analysis, Synthesis, and Realization of the ANALYSIS phase as well as data about future additions and changes of the source databases and their structures. Next, in the Analysis step, we work to understand the future needs of the ideal system based on the information gathered during the Research step. Third, in the Synthesis step, we further synthesize the future needs into possible future scenarios. Finally, in the Realization step, we present the project goal based on the anticipated needs of the data sources and system users. Table 2 presents a summary of this process relative to the PROJECTION phase of the HRADIS project.

The SYNTHESIS phase of the iterative (macro) process design considers how HRADIS will be in a usable state. First, the Research step takes into account the realizations of the previous ANALYSIS and PROJECTION phases to inform the gathering of data on the requirements for the first version of the HRADIS project that addresses its core functions with an understanding of what its future functions will be. Second, the Analysis step involves understanding the requirements of the core functions by creating process and data models for these core functions. Third, the Synthesis step involves the creation of design solutions, and finally, the Realization step involves the development and implementation of these core functionalities. Table 3 presents a summary of this process relative to the SYNTHESIS phase of the HRADIS project.

Table 1. The iterative process model—phase 1 (ANALYSIS).

Macroprocess	Research	Analysis	Synthesis	Realization
ANALYSIS—the true (how it is today)	Gathering data, databases, metadata, and problems researchers face using these data sources	Understanding of the data, data schemas, metadata of each data source, and domains of problems researchers face	Merging the problem domains and data/metadata analysis for the perspective of data warehouse project as a whole	Creating an initial problem statement and requirements document

Table 2. The iterative process model—phase 2 (PROJECTION).

Macroprocess	Research	Analysis	Synthesis	Realization
PROJECTION—the ideal (how it could be)	Gathering additional problems and requirements, including potential data sources to be added and additions and changes to the data and structure of existing data sources	Understanding future data needs and additional requirements of the ideal system	Identifying scenarios that describe user/system interaction of the ideal system from the perspective of the data warehouse project as a whole	Creating use case and project goals documents to include considerations for the future data sources and updates of the existing data sources as well as the requirements of the ideal system

Table 3. The iterative process model—phase 3 (SYNTHESIS).

Macroprocess	Research	Analysis	Synthesis	Realization
SYNTHESIS—the real (how it is tomorrow)	Gathering data on the requirements of the initial version of the data warehouse project that includes its core functions	Understanding of requirements of the core functionalities using process and data modeling tools	Creating design solutions based on process and data models	Development of the design solutions and implementation of the first version of the data warehouse project

The COMMUNICATION phase allows for understanding the process to move the project forward and encompasses the other 3 macroprocess phases of iterative design. The main premise of the COMMUNICATION phase is to keep the entire project team(s) on the same page as the iterative process evolves, and the design continuously changes. Considering a sociotechnical approach, COMMUNICATION also includes how the systems and its users and stakeholders interact.

As its name indicates, the model is both horizontally and vertically iterative (hence the arrows to illustrate the iterative movement); therefore, it allows for continuously updating each shaded box in Figure 3 as we increase our understanding of the requirements and the design of the project.

Results

This section describes the application of the iterative process model to the design research process of problem identification and solution design for HRADIS.

Iterative Process Model—Phase 1: ANALYSIS

As part of the Research step, we downloaded raw data files and data layout and/or data dictionary (metadata) files for all available data releases from the following data sources that are most frequently used by health services researchers: CMS MCR, impact/final rule files, HCAHPS, the area health resources files, AHA annual survey and IT supplement, Dartmouth Atlas, and BLS. Consistent with the iterative process model, the goal was to capture data and metadata from all of the data sources in a single database as is, without changing the source data structure. In general, data and metadata file structures for a given data source were mostly consistent among its releases. Within the release of a particular data source, there were one or multiple data files along with a metadata file. For each data file that was included in the latest release of a particular dataset, we created a table in the data warehouse, naming the table the same as the data file name with a prefix that corresponds to its data source. In cases where previous releases included a data file that was not in the latest release, we also added tables for the additional data files to the data warehouse in the same manner. For example, for the CMS MCR data source, the latest release included 3 data files, namely, ALPHA, NMRC (numeric), and RPT (report). Some of the earlier releases had another data file named ROLLUP. For each of these 4 data files, we created the following tables with an MCR prefix to indicate their data source and a HOSP prefix to distinguish them from other health organization types, for which we may include MCR data in the future: MCR_HOSP_ALPHA, MCR_HOSP_NMRC, MCR_HOSP_RPT, and MCR_HOSP_ROLLUP. Similarly, we created a table named MCR_HOSP_DATAELEMENTS for the metadata files.

Before importing data from data files into the data warehouse, we created an additional column for each table to store the release information. Then, we imported the data values from the data files into their corresponding tables, merging multiple release files into a single corresponding table. When possible, we repeated the same process for the metadata files. The aforementioned processes resulted in a database with data and metadata from all data sources, whose different releases merged into their corresponding tables with their release information preserved. Although data from different sources are not related together as an integrated database yet, having such a database allows for a better understanding of the source data structure and changes to the data structure and metadata over time and also enables data integration from different sources at the query level and the ability to save that query logic. Finally, these manual import processes inform the automation of the import tasks of future releases.

Analysis of the data, metadata, and the problems and issues the health services researchers identified helped inform the Synthesis step. In this step, we considered categories of the problems, each corresponding to a module of the system as a whole (design-focused synthesis and problem categorization). As a result, in the Realization step of the ANALYSIS phase, we generated a list of problems that HRADIS should address, as displayed in Textbox 2.

As part of the Realization step, we also developed the requirements document based on the above problem statement, as displayed in Textbox 3.

Textbox 2. Problem statement at the Realization step of the ANALYSIS phase.

General problems

- Duplication of effort for each project
- Problems related to dealing with a large amount of data
- Management of licenses, data use agreements, and data access levels of users with different roles (administrative, faculty, student, etc)

Integration problems

- Disparate storage of data
- Problems related to dealing with data updates
    - Updates of static data (previous release data does not change; new release data gets added to the previous releases)
    - Updates of dynamic data (new releases add new data; also, update some of the previous release data)
- Integration of data elements from different data sources
    - Lack of standards in how data elements from different sources are integrated
    - Integration of data and metadata

Lack of standards in research data processing to deal with

- Changes of data structure from one data release to another
- Matching data elements from different releases
- Missing data values
- Inconsistent data values
- Variability and lack of documentation of assumptions about the data and the clean-up processes
- Definition and use of measures and indexes

Textbox 3. Requirements document at the Realization step of the ANALYSIS phase.

The system should be able to:

- Store data and metadata from multiple data sources in a single storage (all data should be in one place)
- Store relationships among data elements within and across data sources
- Store rules and procedures for content-specific data processes
    - Imputation of missing values (sometimes even multiple methods for a single data element)
    - Creation of new data elements based on existing ones (calculations, indexes, conversions, etc)
    - Identification of measures and dimensions
- Integrate data sources through the stored relationships, rules, and procedures
- Extract data based on predetermined criteria (data marts)

Iterative Process Model—Phase 2: PROJECTION

Given that the PROJECTION phase deals with *the ideal* during its Research step, collaboration with fellow health services researchers is important to pinpoint potential future problems and requirements and better understand the generic workflow of a hypothetical secondary data analysis research project.

The analysis of the information gathered from the Research step provided insights into the goals and user scenarios for HRADIS. The synthesis of these insights that considers the system as a whole led to the Realization step of the PROJECTION phase, in which we identified a list of tasks and goals for the ideal data warehouse, as displayed in Textbox 4.

Textbox 4. List of tasks and goals for the ideal system at the Realization step of the PROJECTION phase.

- Development of generalized solutions for
  - anticipated data structure changes to the existing data sources
  - addition of new data sources
- User interfaces for the system administrator user role
- In addition to provider-level data, the inclusion of patient-level data
- Addition of data sources about entities health services researchers are interested in other than hospitals (ie, nursing home data)
- Metadata search interface that allows keyword search based on a taxonomy similar to Table 4.
- A user-friendly query builder interface
- An infrastructure that allows
  - multiphase larger projects (harmonious efforts)
  - building new projects based on existing ones
- Inclusion of data analytics toolset
- Inclusion of data visualization toolset
- User interfaces for researchers to utilize analytics and visualization toolsets
- A knowledge base that encompasses metadata, measures and indices, analytics and visualization tools, and references related to all these knowledge base items from the literature

In addition, in the Realization step, we acknowledge the sociotechnical system theory that suggests the development of a technology by always considering the needs of end users. Relative to this project, bringing together various data sources would generate thousands of variables and measures. Moreover, sifting through thousands of variables can be very discouraging unless this process is simplified by considering the needs of the health services researchers. Therefore, to enable seamless development of research projects, an interface that allows intuitive browsing and filtering of metadata through taxonomies is a vital feature of HRADIS. Similar to biologic taxonomy, data taxonomies also separate data elements based on certain common characteristics and simplify browsing [35]. For this purpose, we have developed a data taxonomy (Table 4) by combining our own experience, information on dimensions of health care quality from the CMS Hospital Compare website [36], and hierarchical categories frequently used by health services researchers [37-42]. This taxonomy is incorporated into metadata tables and is dynamic in nature, meaning that one data element can be classified into several categories. In other words, the envisioned user interface will provide some flexibility for health services researchers in categorizing the data elements. This process embodies the use-inspired research model and facilitates further taxonomy growth and development as use and application increase.

Another product of the Realization step of the PROJECTION phase is high-level use cases for the health services researcher and system administrator user roles. Use case analysis is used in systems analysis and design to document the interaction of each user role with the system being considered to be created [43]. Use case analysis is usually performed after requirements definition and user role determination. The use cases are then used for creating the process and data models. For the health services researcher user role, use cases include browsing data elements (metadata) without creating a project, browsing completed projects and selecting one to create a new project by editing it, retrieving data for a given project, and creating a new project. Similarly, for the system administrator user role, use cases include creating system rules, editing system rules, managing user credentials, managing user groups, and managing user and group permissions.

As an example, Figure 4 shows the casual format use case for the researcher user role, which does not include the input/output data elements and their sources/destinations.

The iterative process model allows for considering *the ideal* in early stage design processes. Therefore, decision support tools facilitating the following future state use cases for the health services researcher user role are considered: (1) browsing appropriate data analytics methods, (2) selecting appropriate data analytics methods, (3) browsing data visualization methods, and (4) selecting data visualization methods.

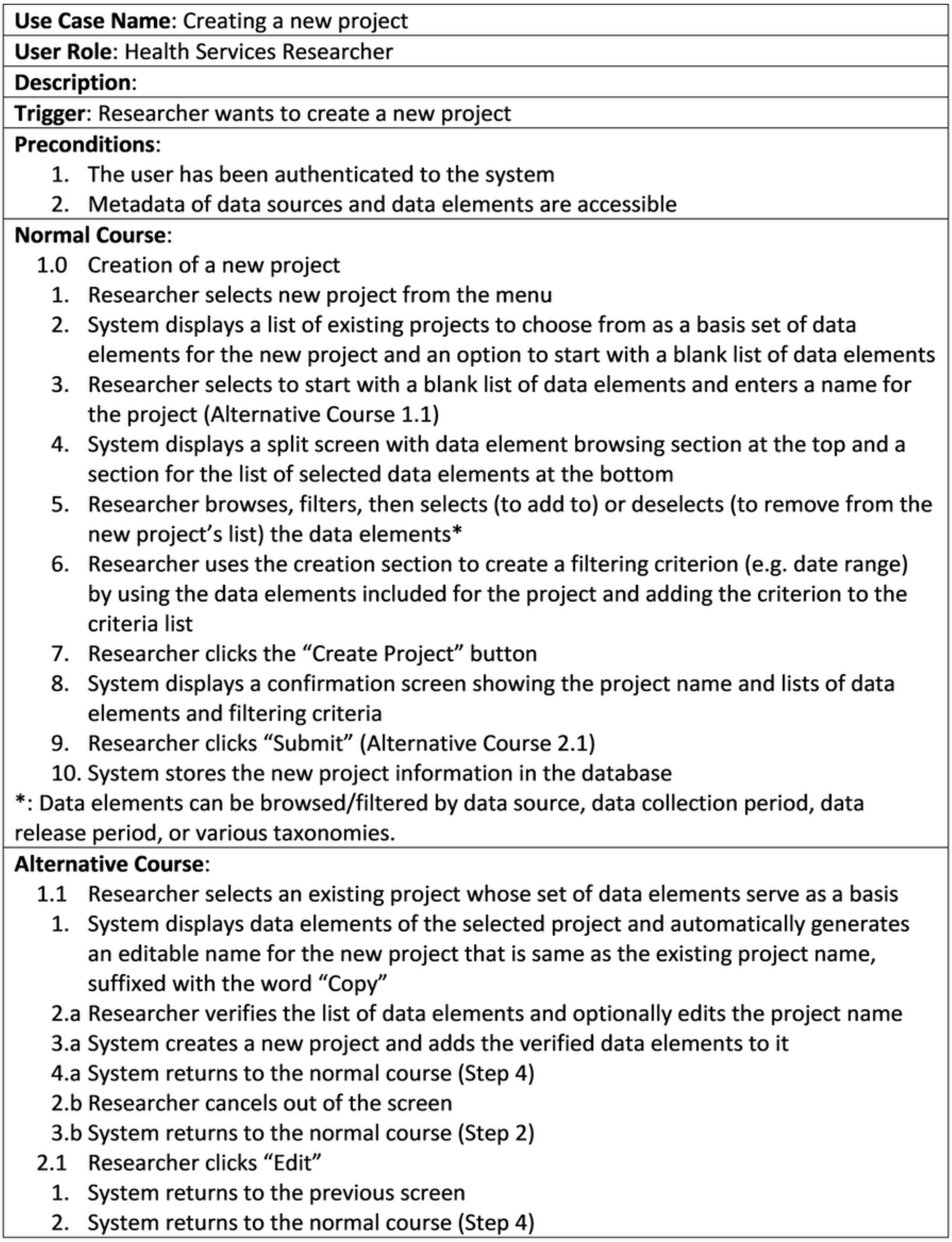
Table 4. Data taxonomy for health services research.

First-level classification	Second-level classification	Examples
Organizational/structural characteristics	• N/A[a]	• Size (number of beds) • Location • System membership
Staffing	• Nurse • Physician • Other	• Registered nurse FTEs[b] per inpatient day • Physician FTEs per inpatient day • Radiology technician staffing
Quality	• Structural measures • Patient experience • Timely and effective care • Outcome measures	• Safe surgery checklist • Communication with doctors • Heart attack—aspirin at arrival • 30-day readmission/mortality
Financial performance	• Profitability • Liquidity • Capital structure • Activity • Utilization	• Operating margin • Current ratio • Equity financing • Total asset turnover • Occupancy rate
Environmental/market characteristics	• N/A	• Market (ie, county, health referral region, or health service area) competition • Managed care penetration • Per capita income (county)

[a] N/A: not applicable.

[b] FTE: full-time equivalent.

Figure 4. A casual format use case example.



Use Case Name: Creating a new project

User Role: Health Services Researcher

Description:

Trigger: Researcher wants to create a new project

Preconditions:

1. The user has been authenticated to the system
2. Metadata of data sources and data elements are accessible

Normal Course:

1.0 Creation of a new project

1. Researcher selects new project from the menu
2. System displays a list of existing projects to choose from as a basis set of data elements for the new project and an option to start with a blank list of data elements
3. Researcher selects to start with a blank list of data elements and enters a name for the project (Alternative Course 1.1)
4. System displays a split screen with data element browsing section at the top and a section for the list of selected data elements at the bottom
5. Researcher browses, filters, then selects (to add to) or deselects (to remove from the new project's list) the data elements*
6. Researcher uses the creation section to create a filtering criterion (e.g. date range) by using the data elements included for the project and adding the criterion to the criteria list
7. Researcher clicks the "Create Project" button
8. System displays a confirmation screen showing the project name and lists of data elements and filtering criteria
9. Researcher clicks "Submit" (Alternative Course 2.1)
10. System stores the new project information in the database

*: Data elements can be browsed/filtered by data source, data collection period, data release period, or various taxonomies.

Alternative Course:

1.1 Researcher selects an existing project whose set of data elements serve as a basis

1. System displays data elements of the selected project and automatically generates an editable name for the new project that is same as the existing project name, suffixed with the word "Copy"

2.a Researcher verifies the list of data elements and optionally edits the project name

3.a System creates a new project and adds the verified data elements to it

4.a System returns to the normal course (Step 4)

2.b Researcher cancels out of the screen

3.b System returns to the normal course (Step 2)

2.1 Researcher clicks "Edit"

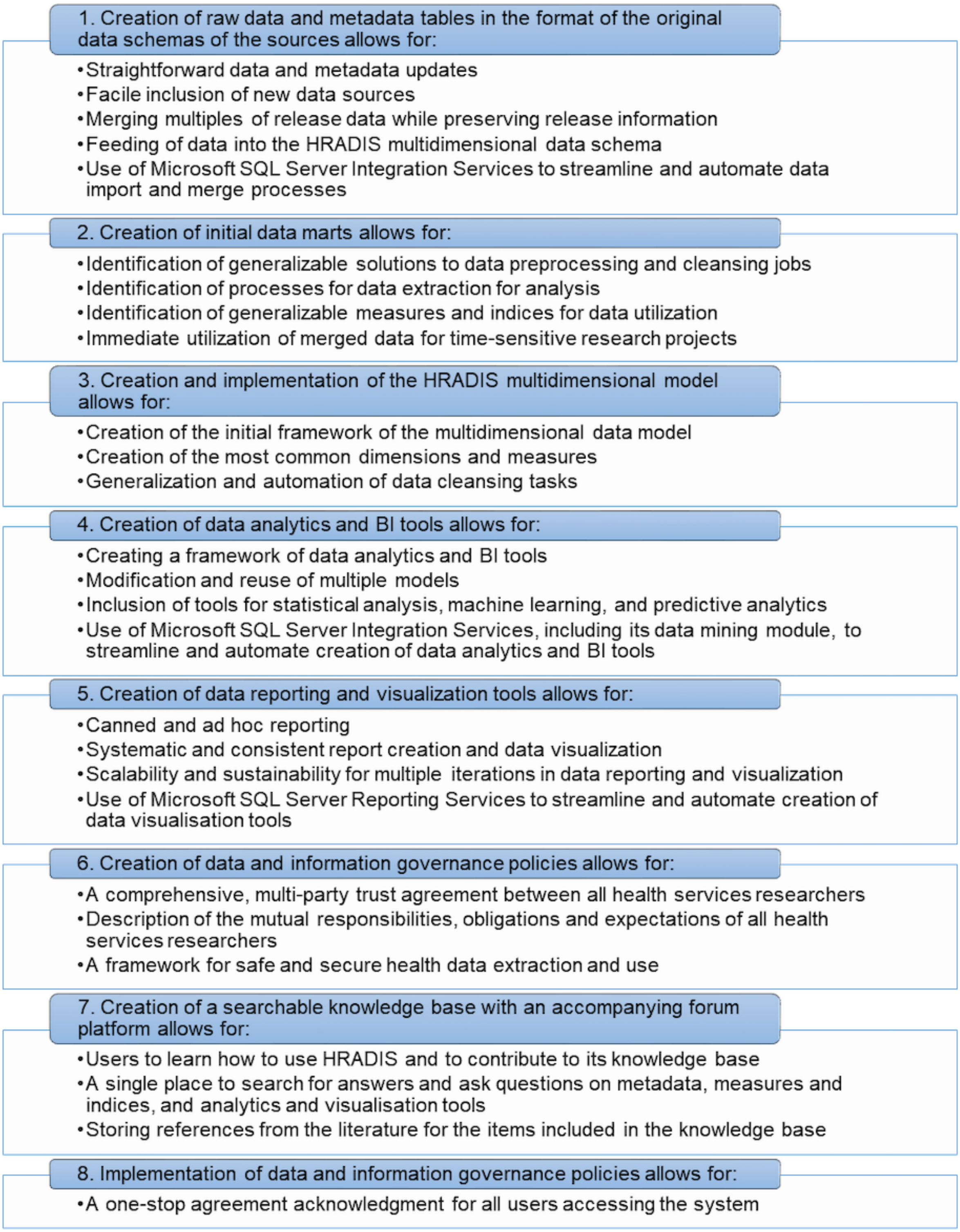
1. System returns to the previous screen
2. System returns to the normal course (Step 4)

Iterative Process Model—Phase 3: SYNTHESIS

Considering that the SYNTHESIS phase focuses on *the real*, its Research step gathers the information from the results of the Realization steps (end products) of the ANALYSIS and PROJECTION phases to develop practical solutions for immediate use. Analysis of the initial problem statement and requirements, use cases, future requirements, and goals of the system revealed that any practical solution has to balance time spent on priority data requests for immediate research projects and time investment required for the development of the data warehouse infrastructure.

For the Realization step, we laid out the process for a plan to implement the first version of HRADIS with its core functionalities. We then expanded this process to include high-level steps to implement some of the future functionalities, such as data analytics and visualization modules. Figure 5 explains the steps of this process, which can also be described as system modules, the reasoning behind why each module is considered, and the tasks involved in each. The iterative process model enables partial-phase completion to develop a system for immediate use, while building out other functionality in an iterative environment. As such, only various parts of the Analysis, Synthesis, and Realization steps of the SYNTHESIS phase were completed. To date, we have implemented the first 3 modules of Figure 5.

Figure 5. The process for a plan to implement healthcare research and analytics data infrastructure solution; BI: business intelligence. HRADIS: healthcare research and analytics data infrastructure solution.



1. Creation of raw data and metadata tables in the format of the original data schemas of the sources allows for:
- Straightforward data and metadata updates
- Facile inclusion of new data sources
- Merging multiples of release data while preserving release information
- Feeding of data into the HRADIS multidimensional data schema
- Use of Microsoft SQL Server Integration Services to streamline and automate data import and merge processes

2. Creation of initial data marts allows for:
- Identification of generalizable solutions to data preprocessing and cleansing jobs
- Identification of processes for data extraction for analysis
- Identification of generalizable measures and indices for data utilization
- Immediate utilization of merged data for time-sensitive research projects

3. Creation and implementation of the HRADIS multidimensional model allows for:
- Creation of the initial framework of the multidimensional data model
- Creation of the most common dimensions and measures
- Generalization and automation of data cleansing tasks

4. Creation of data analytics and BI tools allows for:
- Creating a framework of data analytics and BI tools
- Modification and reuse of multiple models
- Inclusion of tools for statistical analysis, machine learning, and predictive analytics
- Use of Microsoft SQL Server Integration Services, including its data mining module, to streamline and automate creation of data analytics and BI tools

5. Creation of data reporting and visualization tools allows for:
- Canned and ad hoc reporting
- Systematic and consistent report creation and data visualization
- Scalability and sustainability for multiple iterations in data reporting and visualization
- Use of Microsoft SQL Server Reporting Services to streamline and automate creation of data visualisation tools

6. Creation of data and information governance policies allows for:
- A comprehensive, multi-party trust agreement between all health services researchers
- Description of the mutual responsibilities, obligations and expectations of all health services researchers
- A framework for safe and secure health data extraction and use

7. Creation of a searchable knowledge base with an accompanying forum platform allows for:
- Users to learn how to use HRADIS and to contribute to its knowledge base
- A single place to search for answers and ask questions on metadata, measures and indices, and analytics and visualisation tools
- Storing references from the literature for the items included in the knowledge base

8. Implementation of data and information governance policies allows for:
- A one-stop agreement acknowledgment for all users accessing the system

At this stage, as it was essential to obtain a better understanding of the data and the structures in the source systems, we have not yet created a multidimensional data model for HRADIS (a multidimensional model results in faster analysis and output of large and complex datasets). Instead, we kept the data schemas of the source systems in their original state, and table creation only considered the raw data and metadata tables (the first item in Figure 5). These tables will later feed data into the tables based on a new multidimensional HRADIS data schema, whose data model will be designed as part of the Synthesis step (the third item in Figure 5) and will be implemented as part of the Realization step. As part of the process, we merged data from all releases of AHA, Dartmouth Atlas, MCR, Hospital Compare, and HCAHPS data sources into their respective tables in the database. By doing so, we have encountered examples of data preprocessing and cleansing tasks from which to learn and apply in future iterations. As these examples included tasks that are typical for health services research projects, identifying generalized solutions for these assists with isolating the ETL processes, the dimensions and measures for the HRADIS multidimensional model, and the data analytics and visualization tools needed to be created as part of the next steps in the development.

Discussion

Principal Findings

This study provides theoretical underpinnings of the processes and methodologies in developing a data warehouse system as an infrastructure to support health services research. This paper addresses existing inefficiencies, disparate and unnecessary duplication of efforts, and the lack of harmony among health services researchers during the obtaining, extracting, preprocessing, cleansing, and filtering stages of data processing. For this purpose, we envision HRADIS as a full-service data warehouse integrating frequently used health services research data sources, processes, and methods along with a variety of data analytics and visualization tools. A conceptual iterative process model framework combined with sociotechnical systems theory provided guidance on the design process. We presented the application of 4 phases (ie, ANALYSIS, PROJECTION, SYNTHESIS, and COMMUNICATION) of the iterative process model. In the following paragraphs, the discussions on several prominent issues with supporting examples, lessons learned, reflections and recommendations, and future considerations are provided.

In terms of the application of the iterative process model (Figure 3) into the development of HRADIS, the guidance provided to synthesize the ANALYSIS and PROJECTION phases was quite significant. Traditionally, some system development strategies suggest the development and implementation of the core functionalities of a system as the first version, then additional features are added in the later phases or versions. However, the iterative process model provides a solid framework for consideration of the entire system as the pieces are being developed. With the model, the PROJECTION phase guides us to analyze the future requirements of the ideal state of the system before designing the core functionalities of the system for its first version. This allows designers to be informed by the envisioned end product of the PROJECTION phase, hence resulting in the design considerations in the SYNTHESIS phase for the *solutions for the immediate use* to include the goals of the ideal system, some of which will be designed and implemented in the future.

When a small-scale development team comprises only a few members, formal COMMUNICATION may not be as critical. In such an environment, the team is in constant communication naturally and is able to utilize agile development methods, where the features that are immediately needed are analyzed, designed, and implemented. The implemented features satisfy the immediate requirements and may later go through slight modifications to be generalized and fit into the larger project. This is also how the HRADIS project was initially implemented, by creating ad hoc data extracts, transform and load procedures, and queries for immediate research projects. Working with larger teams, on the other hand, requires more formal COMMUNICATION to create a shared understanding of the immediate processes as well as awareness of the larger *to be* project. As mentioned earlier, how the system communicates with the users and stakeholders in the general sense, and COMMUNICATION in this specific context, is critical, for example, interactions between the system and its users when there is a request for a new data source to be included in the warehouse as well as when a data extract is requested from the system.

Conclusions

During the development of HRADIS, several issues were found that are worth further discussion. We believe that for those who consider attempting a similar project, the following lessons learned, reflections, and recommendations would be instrumental.

First, seamless progress requires a balance between immediate/urgent needs and the need to generalize the solutions being considered. To achieve this balance, we developed practical solutions by recognizing the trade-off between the quality and cost during the SYNTHESIS phase. The time investment into a highly generalized, reusable, better-quality solution to a specific problem that would yield time savings in the long term comes with its opportunity cost of not spending that particular time into multiple, less effective but working, ad hoc solutions that may yield results in the short term. For example, as we considered various geographical categorizations of hospitals based on county, health service area (HSA), and health referral region (HRR) codes in the AHA data for a study, we recognized missing values in the data and considered several ways to calculate the missing codes. We were faced with making a decision between creating a generalized solution that would encompass all possible ways to calculate the missing values or create an ad hoc solution specific to the pilot project we were working on at the time. The generalized solution would take a longer time investment to create, with the potential to be used for many studies, compared with the specific solution. In this particular case, we chose to implement the generalized solution as the long-term benefits of reuse outweighed the opportunity cost of delaying the use of HRADIS for particular pilot projects. However, these types of decisions must be considered on a case-by-case basis, as the decision would strongly depend on the potential reuse of the generalized solution and the urgency of the particular study. When making decisions on such trade-offs, one should also consider the potential benefit of the ad hoc solution in developing a generalized solution given that the ad hoc solutions sometimes provide the required knowledge base and intimacy between the designer and the data.

Second, the issue of static versus dynamic data import that was mentioned in Textbox 2 is an important consideration. When developing general solutions for data import, we realized that there was a need for two different approaches for data import processes. This need was due to inherent differences in the data sources. The former approach is static as the data source itself is static, meaning that once data are published, the content of the data does not change over time. The latter approach is dynamic as the data source itself is dynamic, meaning that the data are updated at regular intervals, and the content changes even for the archived versions (ie, years) of the data. A good example of a dynamic data source is CMS MCR; reports for earlier years can be reopened after settlement, and even the archived data are updated quarterly [44]. In our case, importing data from static data sources did not require much effort, as it was sufficient to create simple SQL scripts for import tasks. Importing data from dynamic sources requires writing SQL stored procedures that automatize and simplify the quarterly data import processes. This process addressed our ultimate goal to improve research efficiency and reduce the amount of time spent on redundant tasks.

Third, as mentioned in Textbox 3, when designing such a data warehouse, the team may consider potential ways to improve data by utilizing different data sources. In our case, sometimes, the same variable or measure existed in different datasets or was sourced from another dataset. To enhance the completeness of the data and address any missing value issues, we examined both datasets by comparing and ultimately imputing the missing values. For example, when developing certain measures, such as the Herfindahl-Hirshman index, we needed to use certain geographical market area designations such as HSA, county, or HRRs. However, due to missing information in the existing dataset for certain years, we realized that there is a need to examine the original data source (Dartmouth Atlas). Further examination revealed that the missing information could be imputed by developing an algorithm that utilizes both the information from the original data source and the existing dataset.

The fourth lesson learned pertains to the importance of the iterative design process. The conceptual iterative process model framework adapted from Jonas [34] was very useful during the development of HRADIS. Although the iterative back and forth movements may be initially perceived as inefficient and time-consuming, they were crucial in developing generalized design solutions that are beneficial in the long term. Although it may be tempting to develop a system in response to urgent data needs, we found it essential to adhere to the iterative process model. Doing so created a development expectation with our colleagues.

In the future, we plan to improve HRADIS by drawing on by the successful growth strategy and story of research electronic data capture (REDCap). Doing so considers that both HRADIS and REDCap are products of academic research and have ambitious goals, but they start small because of limited resources [45]. We plan to collaborate with researchers who have potential contributions by asking them to work with us in generalizing their contributions to fit the HRADIS framework. In this way, the contributor would have access to all the HRADIS offerings, and the existing user base would have access to the new contribution (within the data governance limitations).

Conflicts of Interest

None declared.

References

1. Agency for Healthcare Research and Quality. 2015. An Organizational Guide to Building Health Services Research Capacity URL: https://www.ahrq.gov/funding/training-grants/hsrguide/hsrguide.html [accessed 2020-01-23]
2. Puppala M, He T, Chen S, Ogunti R, Yu X, Li F, et al. METEOR: an enterprise health informatics environment to support evidence-based medicine. IEEE Trans Biomed Eng 2015 Dec;62(12):2776-2786. [doi: 10.1109/TBME.2015.2450181] [Medline: 26126271]
3. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood) 2014 Jul;33(7):1115-1122. [doi: 10.1377/hlthaff.2014.0147] [Medline: 25006136]
4. Ozaydin B, Houser S, Zengul F. Tracking the journey of electronic health data through health information systems. J AHIMA 2018 Jan;1(18):40-44 [FREE Full text]
5. Berndt DJ, Hevner AR, Studnicki J. The catch data warehouse: support for community health care decision-making. Decis Support Syst 2003 Jun;35(3):367-384. [doi: 10.1016/s0167-9236(02)00114-8]
6. Tremblay MC, Fuller R, Berndt D, Studnicki J. Doing more with more information: changing healthcare planning with OLAP tools. Decis Support Syst 2007 Aug;43(4):1305-1320. [doi: 10.1016/j.dss.2006.02.008]
7. ResDAC: Research Data Assistance Center, University of Minnesota. 2020. Our Services URL: https://www.resdac.org/about-resdac/our-services [accessed 2020-01-23]
8. Wharton Research Data Services. 2020. URL: http://www.whartonwrds.com/ [accessed 2020-01-23]
9. Kroenke D, Auer D. Database Processing: Fundamentals, Design, and Implementation. New York, USA: Pearson; 2015.
10. NBER: The National Bureau of Economic Research. 2020. Healthcare Cost Report Information System (HCRIS) Data URL: http://www.nber.org/data/hcris-hosp.html [accessed 2020-01-23]
11. The American Hospital Directory. 2020. URL: https://www.ahd.com/ [accessed 2020-01-23]
12. Hospital-Data. 2020. Hospital and Nursing Home Profiles URL: http://www.hospital-data.com/ [accessed 2020-01-23]

13. Manuel Pérez-Martínez J, Berlanga-Llavori R, Aramburu-Cabo MJ, Pedersen TB. Contextualizing data warehouses with documents. Decis Support Syst 2008 Apr;45(1):77-94. [doi: 10.1016/j.dss.2006.12.005]
14. Corral K, Schuff D, St Louis RD. The impact of alternative diagrams on the accuracy of recall: a comparison of star-schema diagrams and entity-relationship diagrams. Decis Support Syst 2006 Oct;42(1):450-468. [doi: 10.1016/j.dss.2005.02.003]
15. Schuff D, Corral K, Turetken O. Comparing the understandability of alternative data warehouse schemas: an empirical study. Decis Support Syst 2011 Dec;52(1):9-20. [doi: 10.1016/j.dss.2011.04.003]
16. Trist EL, Bamforth KW. Some social and psychological consequences of the longwall method of coal-getting. Hum Relat 2016 Apr 22;4(1):3-38. [doi: 10.1177/001872675100400101]
17. Cherns A. The principles of sociotechnical design. Hum Relat 2016 Apr 22;29(8):783-792. [doi: 10.1177/001872677602900806]
18. Ulhøi J, Jørgensen F. Linking humanity with performability through social-technical systems theory. Int J Perform Eng 2010;6(1):89-99 [FREE Full text]
19. Organ J, Stapleton L. Technologist engagement with risk management practices during systems development? Approaches, effectiveness and challenges. AI Soc 2015 Aug 1;31(3):347-359. [doi: 10.1007/s00146-015-0597-4]
20. Appelbaum SH. Socio ‐ technical systems theory: an intervention strategy for organizational development. Manag Decis 1997 Aug;35(6):452-463. [doi: 10.1108/00251749710173823]
21. Geels FW. From sectoral systems of innovation to socio-technical systems. Res Pol 2004 Sep;33(6-7):897-920. [doi: 10.1016/j.respol.2004.01.015]
22. Eason K. Before the internet: the relevance of socio-technical systems theory to emerging forms of virtual organisation. Int J Sociotechnology Knowl Dev 2009;1(2):23-32 [FREE Full text] [doi: 10.4018/jskd.2009040103]
23. Bélanger F, Watson-Manheim MB, Swan BR. A multi-level socio-technical systems telecommuting framework. Behav Inform Technol 2013 Dec;32(12):1257-1279. [doi: 10.1080/0144929x.2012.705894]
24. Wang J, Solan D, Ghods A. Distance learning success – a perspective from socio-technical systems theory. Behav Inform Technol 2010 May;29(3):321-329. [doi: 10.1080/01449290903544645]
25. Oriol M, Tumulty G. IGI Global. 2009. Successful Online Teaching and Learning Strategies URL: https://www.igi-global.com/chapter/successful-online-teaching-learning-strategies/27326 [accessed 2020-05-15]
26. Kim HN. The phenomenon of blogs and theoretical model of blog use in educational contexts. Comput Educ 2008 Nov;51(3):1342-1352. [doi: 10.1016/j.compedu.2007.12.005]
27. Shin D, Choi MJ. Ecological views of big data: perspectives and issues. Telematics Inf 2015 May;32(2):311-320. [doi: 10.1016/j.tele.2014.09.006]
28. Jones D, Gregor S. The anatomy of a design theory. J Assoc Inf Syst 2007 May;8(5):312-335. [doi: 10.17705/1jais.00129]
29. Hevner A, Chatterjee S. Design Research in Information Systems: Theory and Practice. New York, USA: Springer; 2010.
30. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q 2004;28(1):75. [doi: 10.2307/25148625]
31. Offermann P, Levina O, Schönherr M, Bub U. Outline of a Design Science Research Process. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. 2009 Presented at: DESRIST'09; May 6-8, 2019; Philadelphia, USA. [doi: 10.1145/1555619.1555629]
32. Giorgini P, Rizzi S, Garzetti M. GRAnD: a goal-oriented approach to requirement analysis in data warehouses. Decis Support Syst 2008 Apr;45(1):4-21. [doi: 10.1016/j.dss.2006.12.001]
33. Rosenkranz C, Holten R, Räkers M, Behrmann W. Supporting the design of data integration requirements during the development of data warehouses: a communication theory-based approach. Eur J Inform Syst 2017 Dec 19;26(1):84-115. [doi: 10.1057/ejis.2015.22]
34. Jonas W. Research through DESIGN through research. Kybernetes 2007 Oct 23;36(9/10):1362-1380. [doi: 10.1108/03684920710827355]
35. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
36. Medicare: Hospital Compare. 2020. Measures and Current Data Collection Periods URL: https://www.medicare.gov/hospitalcompare/Data/Data-Updated.html# [accessed 2020-01-23]
37. Everhart D, Neff D, Al-Amin M, Nogle J, Weech-Maldonado R. The effects of nurse staffing on hospital financial performance: competitive versus less competitive markets. Health Care Manage Rev 2013;38(2):146-155 [FREE Full text] [doi: 10.1097/HMR.0b013e318257292b] [Medline: 22543824]
38. Harless DW, Mark BA. Nurse staffing and quality of care with direct measurement of inpatient staffing. Med Care 2010 Jul;48(7):659-663. [doi: 10.1097/MLR.0b013e3181dbe200] [Medline: 20548254]
39. Mark BA, Harless DW, McCue M. The impact of HMO penetration on the relationship between nurse staffing and quality. Health Econ 2005 Jul;14(7):737-753. [doi: 10.1002/hec.988] [Medline: 15744748]
40. Mark BA, Harless DW, McCue M, Xu Y. A longitudinal examination of hospital registered nurse staffing and quality of care. Health Serv Res 2004 Apr;39(2):279-300 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00228.x] [Medline: 15032955]

41. Pink GH, Holmes GM, D'Alpe C, Strunk LA, McGee P, Slifkin RT. Financial indicators for critical access hospitals. J Rural Health 2006;22(3):229-236. [doi: 10.1111/j.1748-0361.2006.00037.x] [Medline: 16824167]
42. Zengul FD, Weech-Maldonado R, Ozaydin B, Patrician PA, O Connor SJ. Longitudinal analysis of high-technology medical services and hospital financial performance. Health Care Manage Rev 2018;43(1):2-11. [doi: 10.1097/HMR.0000000000000124] [Medline: 27467169]
43. Dennis A, Wixom B, Roth R. Systems Analysis and Design. New Jersey, USA: Wiley; 2014.
44. Asper F. ResDAC. 2013. Understanding the Data URL: http://resdac.umn.edu/sites/resdac.umn.edu/files/Understanding%20the%20Data%20(Slides).pdf [accessed 2020-01-23]
45. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

Abbreviations

AHA: American Hospital Association
BI: business intelligence
BLS: Bureau of Labor Statistics
CMS: Centers for Medicare and Medicaid Services
EHR: electronic health record
ETL: extract, transform, and load
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
HRADIS: Healthcare Research and Analytics Data Infrastructure Solution
HRR: health referral region
HSA: health service area
IT: information technology
MCR: Medicare cost reports
NBER: National Bureau of Economic Research
ResDAC: Research Data Assistance Center
REDCap: research electronic data capture
WRDS: Wharton research data services

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by J Lee, M Alshawmar; comments to author 23.03.20; revised version received 08.04.20; accepted 16.04.20; published 04.06.20.

Please cite as:
Ozaydin B, Zengul F, Oner N, Feldman SS
Healthcare Research and Analytics Data Infrastructure Solution: A Data Warehouse for Health Services Research
J Med Internet Res 2020;22(6):e18579
URL: https://www.jmir.org/2020/6/e18579
doi:10.2196/18579
PMID:32496199

Original Paper

Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach

Cheng-Sheng Yu[1,2], PhD; Yu-Jiun Lin[1,2], MD; Chang-Hsien Lin[1,2], MD; Shiyng-Yu Lin[1,2], MD; Jenny L Wu[1,2], BSc; Shy-Shin Chang[1,2], MD, PhD

[1]Department of Family Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[2]Department of Family Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Shy-Shin Chang, MD, PhD
Department of Family Medicine
School of Medicine
College of Medicine, Taipei Medical University
250 Wuxing St
Taipei, 11031
Taiwan
Phone: 886 2 23565926
Email: sschang0529@gmail.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/7/e21753/

Abstract

Background: In the era of information explosion, the use of the internet to assist with clinical practice and diagnosis has become a cutting-edge area of research. The application of medical informatics allows patients to be aware of their clinical conditions, which may contribute toward the prevention of several chronic diseases and disorders.

Objective: In this study, we applied machine learning techniques to construct a medical database system from electronic medical records (EMRs) of subjects who have undergone health examination. This system aims to provide online self-health evaluation to clinicians and patients worldwide, enabling personalized health and preventive health.

Methods: We built a medical database system based on the literature, and data preprocessing and cleaning were performed for the database. We utilized both supervised and unsupervised machine learning technology to analyze the EMR data to establish prediction models. The models with EMR databases were then applied to the internet platform.

Results: The validation data were used to validate the online diagnosis prediction system. The accuracy of the prediction model for metabolic syndrome reached 91%, and the area under the receiver operating characteristic (ROC) curve was 0.904 in this system. For chronic kidney disease, the prediction accuracy of the model reached 94.7%, and the area under the ROC curve (AUC) was 0.982. In addition, the system also provided disease diagnosis visualization via clustering, allowing users to check their outcome compared with those in the medical database, enabling increased awareness for a healthier lifestyle.

Conclusions: Our web-based health care machine learning system allowed users to access online diagnosis predictions and provided a health examination report. Users could understand and review their health status accordingly. In the future, we aim to connect hospitals worldwide with our platform, so that health care practitioners can make diagnoses or provide patient education to remote patients. This platform can increase the value of preventive medicine and telemedicine.

(J Med Internet Res 2020;22(6):e18585) doi:10.2196/18585

KEYWORDS

machine learning; online healthcare assessment; medical informatics; preventive medicine

Introduction

In the ever-changing technological era, the internet can provide rapid and convenient medical services in the form of health care, preventive medicine, and telemedicine. Medical informatics is a multidisciplinary field that comprises medicine and computer science. As computer technology continues to advance, medical informatics can be used to develop various applications such as electronic medical records (EMRs), medical image processing, clinical diagnosis decision systems, hospital information management systems, telemedicine, and internet and health information systems [1-4].

To construct a health care information system, several factors must be considered: the hospital information system, including both clinical management and diagnosis services; the storage and processing of patient information, such as EMRs and electronic health records; decision support systems, such as expert diagnosis systems; and the artificial intelligence (AI) algorithms that need to be applied to those factors (eg, data mining in EMRs and decision-making in clinical diagnosis) [5-8].

The mass application of EMRs and the digitalization of medical equipment and instruments have led to the continuous expansion of information capacity in hospital databases. Therefore, informatics research should focus on basic electronic medical database construction, data collection and analysis, medical decision support, and automatic knowledge acquisition. Furthermore, the use of machine learning (ML) technology in AI to extract the most important information has led to cutting-edge research in medicine [9-13]. The goal of AI is to construct an intelligent machine that imitates the natural intelligence of humans. Computers, robots, and software that are made with such technology will have human-like thinking processes, but with the ability to utilize superhuman speed and power effectively. Knowledge engineering is an essential part of AI research, especially ML, because AI operations require a significant amount of real-world data.

ML is defined as a “machine that is capable of self-learning without any guidance.” Therefore, the main purpose of ML is to make computers self-learning and auto-correcting when analyzing data. The core technology of ML must identify specific patterns and information hidden within very large data sets using statistical analysis and prediction automatically [14-17].

Disease and disability are influenced by several factors: environmental factors, genetic predisposition, pathogens, and lifestyle choices. Some conditions are a dynamic process that can affect an individual before they are aware of any problem [18-20]. The core of preventive medicine is to prevent chronic diseases among people who are at risk of certain diseases. In some cases, it can also be used to reverse their condition, returning them to a good health status. In the past, due to information asymmetry, doctors and hospitals led the medical environment, and patients did not have access to any appropriate methods or information to implement real-time self-management. Patients who failed to obtain an early diagnosis would have to pay higher health care costs. Therefore, the spirit of prevention medicine is that “an ounce of prevention is worth a pound of cure” [21-23].

Metabolic syndrome (MetS) is a cluster of conditions comprising high blood sugar, high blood pressure, abnormal blood lipid levels, abdominal obesity, and other metabolic risk factors. It is a warning sign of potential future chronic disease. People with MetS have an increased risk of subsequent development of type II diabetes, hypertension, hyperlipidemia, heart disease, and stroke compared with healthy people [24-28].

Chronic kidney disease (CKD) is defined as kidney function that is impaired for longer than 3 months, leading to irreversible damage. The National Kidney Foundation Kidney Disease Outcome Quality Initiative guideline classifies CKD into 5 stages according to the estimated glomerular filtration rate (eGFR) and using the recommended Modification of Diet in Renal Disease (MDRD) equation [29]. There are many causes of CKD, such as congenital anomalies of the kidney, urinary tract obstruction, urinary tract infection, and glomerulopathy. In addition, hypertension, diabetes, and gout are common chronic diseases that cause CKD if undertreated [30,31].

Telemedicine uses information and telecommunication technology to deliver medical information and physicians’ diagnoses to patients without the limitations of time and space. It combines information and communication technologies with medical expertise to provide various services: remote consultation and conferencing for doctors; comprehensive medical care for residents in remote and outlying islands; and teaching and training opportunities for medical staff. The internet can be used to assist with the popularization of telemedicine to achieve a two-way communication channel between patients and medical practitioners [32,33]. Therefore, this study aims to construct an online ML-driven medical database system from EMRs of subjects who have undergone health examination, and provide online self-health evaluation for MetS and CKD.

Methods

Setting

The study was conducted at the Health Management Center (HMC) of Taipei Medical University Hospital (TMUH). Electronic medical records (EMRs) were obtained and reviewed from the HMC, which receives approximately 60 to 70 visits per month.

Ethics

The study was approved by the Institutional Review Board (IRB) of TMUH prior to data collection (TMUH TMU-JIRB number N202003088), in accordance with the original and amended Declaration of Helsinki. The IRB waived the need for informed consent because of the retrospective nature of this study.

EMR Database and System

The databases and the selected predicting variables (Table 1) were derived from previous publication on MetS and CKD [34-36]. Figure 1 shows an overview of the system and the main functions. Briefly, using a series of complicated procedures, the two databases (MetS and CKD) were connected to an internet platform to construct one integrated system. This web-based system was embedded with ML models to provide various medical evaluations and analyses. The online system was constructed on a server as a web-based environment. The frontend implementation included the programming language JavaScript (Oracle Corp), the framework VueJS (Vue), and the styling Syntactically Awesome Style Sheets (Sass). The backend implementation used Java and R as the programming languages, and all ML calculations and evaluations were conducted using the statistical program R (version 3.6.1, R Foundation for Statistical Computing). The back web framework was Spring Boot (Pivotal Software), connecting the MySQL (Oracle Corp) database as the storage system.

Study Populations

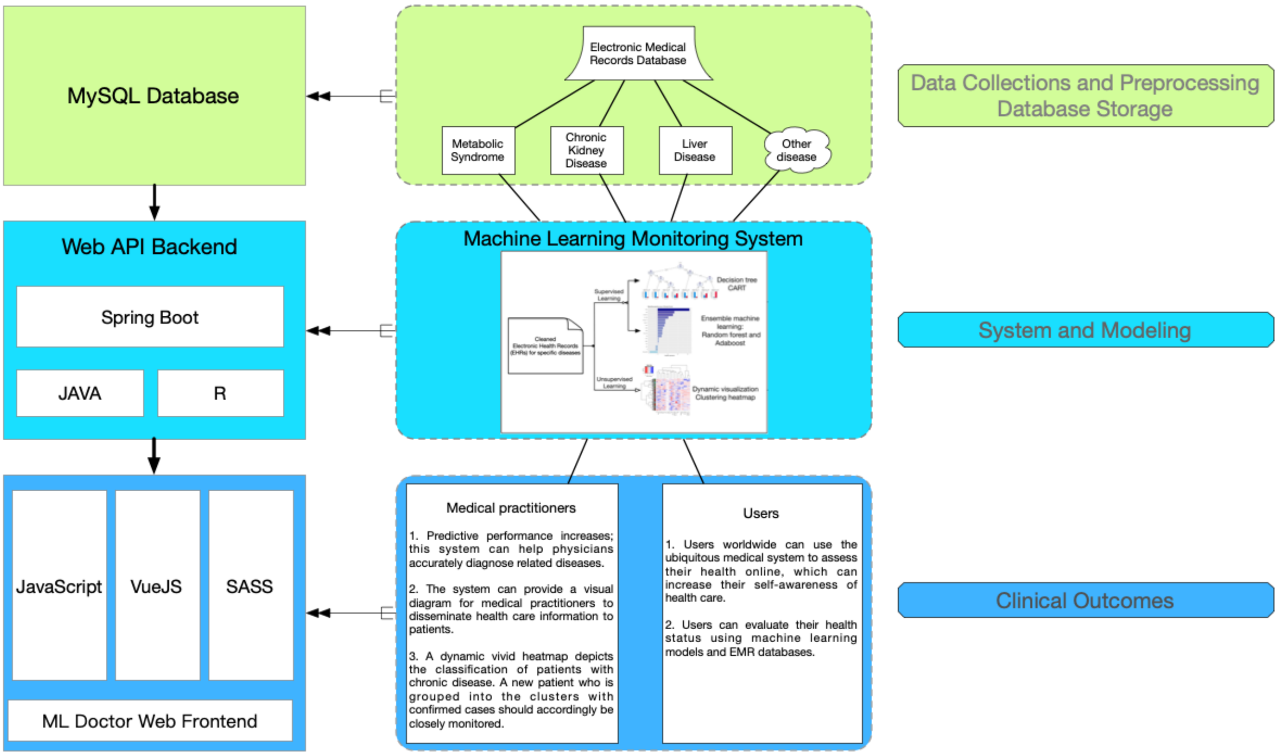
Figure 2 [37-39] shows an overview of the main study population and the validation populations. Briefly, the starting study population included 48,628 EMRs of Taiwanese adults aged over 18 years who underwent a self-paid health examination at TMUH from July 2015 to December 2019. All the study participants completed a self-questionnaire on demographics, existing medical conditions, and the use of medications.

Table 1. The list of predicting variables in the electronic health care records.

Disease and predicting variable	Unit
Metabolic syndrome	
Sex	Male/Female
Age	years
Body mass index	kg/m^2
Waist circumference	cm
Glutamic-oxaloacetic transaminase	IU/L
Glutamate pyruvate transaminase	IU/L
γ-Glutamyl transpeptidase	U/L
Total bilirubin	mg/dL
Alkaline phosphatase	IU/L
Blood urea nitrogen	mg/dL
Creatinine	mg/dL
Uric acid	mg/dL
Albumin	g/dL
Cholesterol	mg/dL
High-density lipoprotein	mg/dL
Low-density lipoprotein	mg/dL
Hemoglobin A_{1c}	%
Glucose AC	mg/dL
Triglycerides	mg/dL
Systolic blood pressure	mm Hg
Diastolic blood pressure	mm Hg
Elastic modulus (E) score	kPa
Controlled attenuation parameter (CAP) score	dB/m
Chronic kidney disease	
Sex	Male/Female
Age	years
Body mass index	kg/m^2
Waist circumference	cm
Glutamic-oxaloacetic transaminase	IU/L
Glutamate pyruvate transaminase	IU/L
γ-Glutamyl transpeptidase	U/L
Total bilirubin	mg/dL
Alkaline phosphatase	IU/L
Blood urea nitrogen	mg/dL
Creatinine	mg/dL
Uric acid	mg/dL
Albumin	g/dL
Cholesterol	mg/dL
High-density lipoprotein	mg/dL
Low-density lipoprotein	mg/dL

Disease and predicting variable	Unit
Hemoglobin A_{1c}	%
Hypertension	Yes/No

Figure 1. The structure of web-based machine learning medical system. API: application programming interface; EMR: electronic medical record; ML: machine learning.



Medical practitioners

1. Predictive performance increases; this system can help physicians accurately diagnose related diseases.

2. The system can provide a visual diagram for medical practitioners to disseminate health care information to patients.

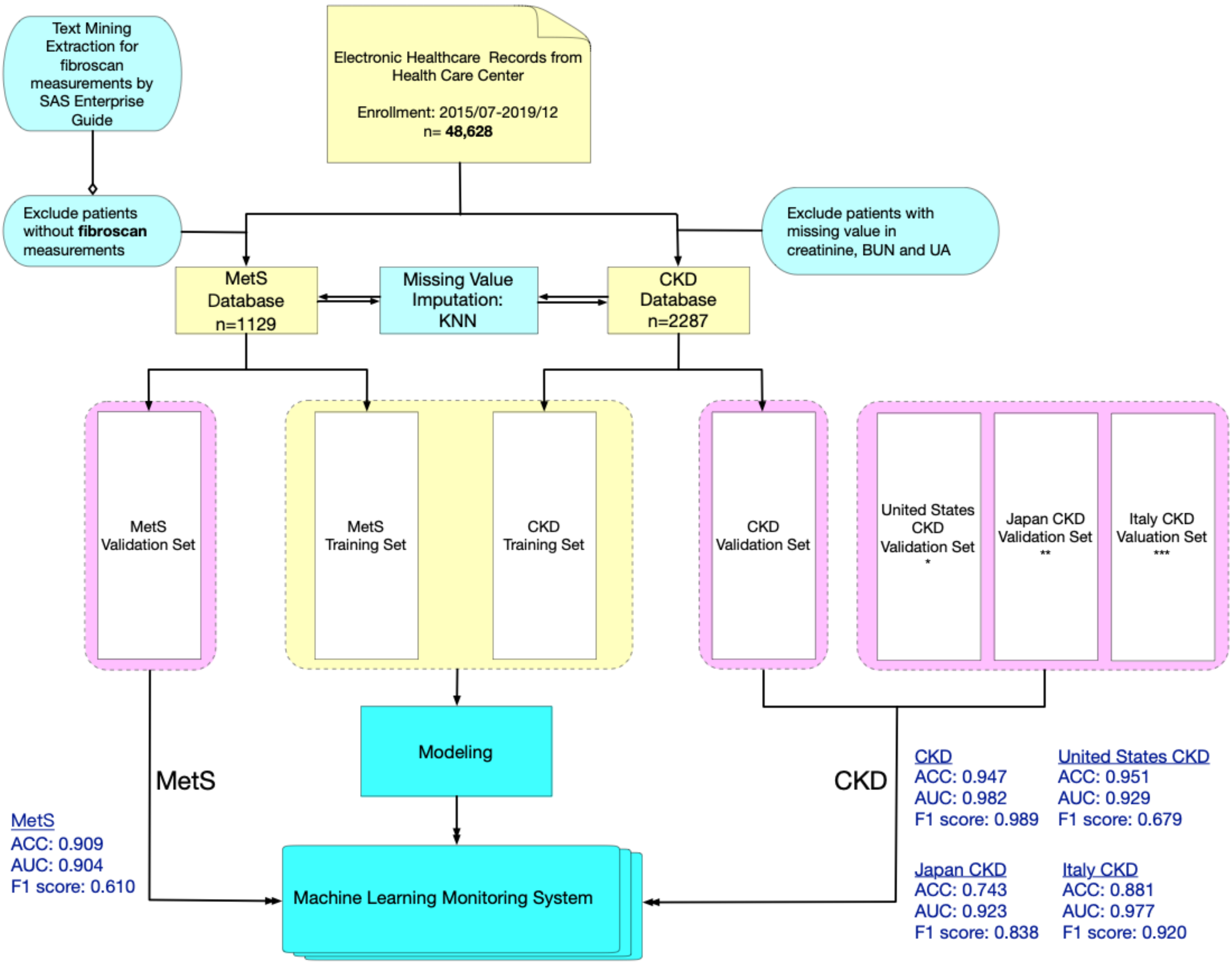
3. A dynamic vivid heatmap depicts the classification of patients with chronic disease. A new patient who is grouped into the clusters with confirmed cases should accordingly be closely monitored.

Users

1. Users worldwide can use the ubiquitous medical system to assess their health online, which can increase their self-awareness of health care.

2. Users can evaluate their health status using machine learning models and EMR databases.

Figure 2. Flowchart of data collection and preprocessing for MetS and CKD data sets including training and validation sets. SAS Enterprise Guide is a software that combines the analytic ability of SAS software with a user-friendly interface. It provides several functions of Structured Query Language (SQL), which includes a text mining technique. ACC: accuracy; AUC: area under the curve; BUN: blood urea nitrogen; CKD: chronic kidney disease; KNN: k-nearest neighbors algorithm; MetS: metabolic syndrome; UA: uric acid. * Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) [37], ** Iimori et al [39], *** De Nicola et al [38].



Subsequently, the starting population data underwent data cleaning and preprocessing to form two distinct databases (MetS and CKD) for ML. For the MetS database, there were a total of 1129 participants after the exclusion of participants without FibroScan (Echosens) measurements. For the CKD database, there were a total of 2287 participants after the exclusion of participants without values for creatinine, blood urea nitrogen, and uric acid.

Due to the inconsistent definition of MetS across the world, the ML performance of the MetS database and the CKD database were validated using different study populations. The ML performance of the CKD database was validated using Taiwanese, Italian, US, and Japanese data sets, but the ML performance of the CKD database was only validated using a Taiwanese data set [37-39]. Since different variables may be unavailable in different validation data sets, unavailable variables were simply excluded in ML performance analysis for a balanced comparison.

ML Techniques

The ML techniques used in this system included supervised learning models, such as classification and regression tree (CART) and random forest [35,36]. Supervised learning was applied to classify the patients in the training set and predict patients with a specific chronic disease or syndrome in the validation set before the prediction model was available on this system [40,41]. In addition, unsupervised learning (hierarchical clustering using the Ward method and Euclidean distance) was embedded in a heat map, providing classified visualization between new input records and the database. An interactive heat map that could be rearranged or zoomed in and out was applied to this system [42-47].

All outcomes were presented on the web platform after the ML system evaluated the users' EMRs. Although the ML system was developed on a web-based interface, it could be embedded in the Internet of Medical Things (IoMT) environment, for example, as apps or real-time monitoring systems between several medical centers and hospitals [48-50].

Questionnaire Selection

To measure the usability of websites, we invited potential users of the ML system (physicians, medical staff, and potential users) to fill out a system usability scale (SUS) evaluation questionnaire. SUS was chosen as the usability test tool because previous studies found it to be reliable and quick to answer, and the final score is provided with interpretation based on a well-established reference standard [51,52]. In general, the higher the SUS score, the better the usability of the website. Details about the questionnaire design (the 10 questions), score summary, and results of reliability and validity tests are given in Multimedia Appendix 1.

Results

The web-based health care ML system provides online diagnosis of three diseases (Figure 3), and it is available on the internet [53]. The website provides an assessment of MetS and CKD; the system for noncancer liver disease is still under beta testing. Report pages are provided for online diagnosis of each disease. Therefore, users from all over the world can choose the evaluation provided depending on their requirements. Users input the predicting variables (Table 1) into the website to evaluate their health (Figure 4), and the evaluation results will appear in <5 seconds when there is a single request. Missing predicting variables are allowed, and the missing values will be imputed based on the mean values from the database. However, the users are warned that missing predicting variables may result in poorer prediction accuracy. The details of stress tests with different numbers of requests (100 to 800) can be found in Multimedia Appendix 2. Briefly, a stress test with 800 requests reports a throughput of 4.7 requests per second. To evaluate the usability of the system, we invited 30 volunteers to complete the SUS evaluation questionnaire. The volunteers included 6 physicians, 12 medical staff, and 12 potential users (Multimedia Appendix 1). It was found that the average SUS score is 74, which indicates a good usability rating [54]. In addition, results were found to be reliable and valid by Kaiser-Meyer-Olkin and Bartlett tests. The entire analysis process follows a strict privacy policy, so that none of the patients' private information is ever recorded.

Figure 3. Home page of the machine learning health care system.

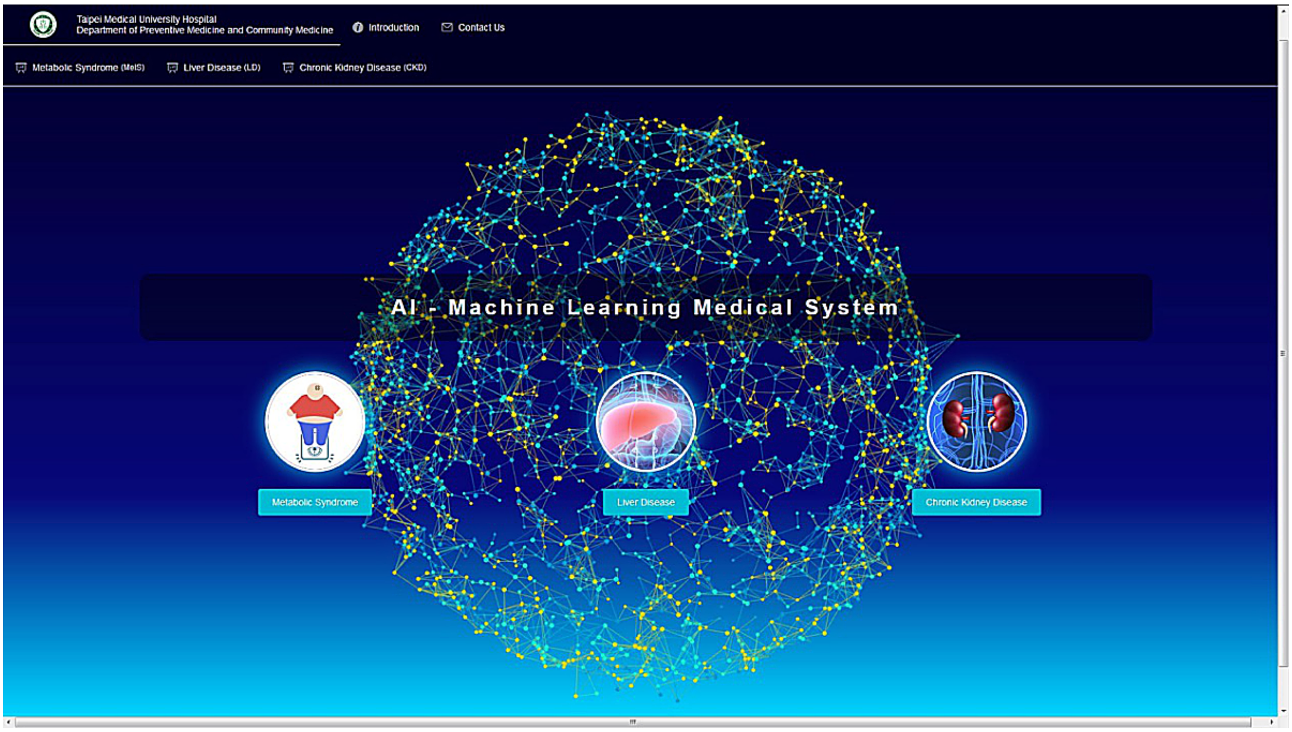
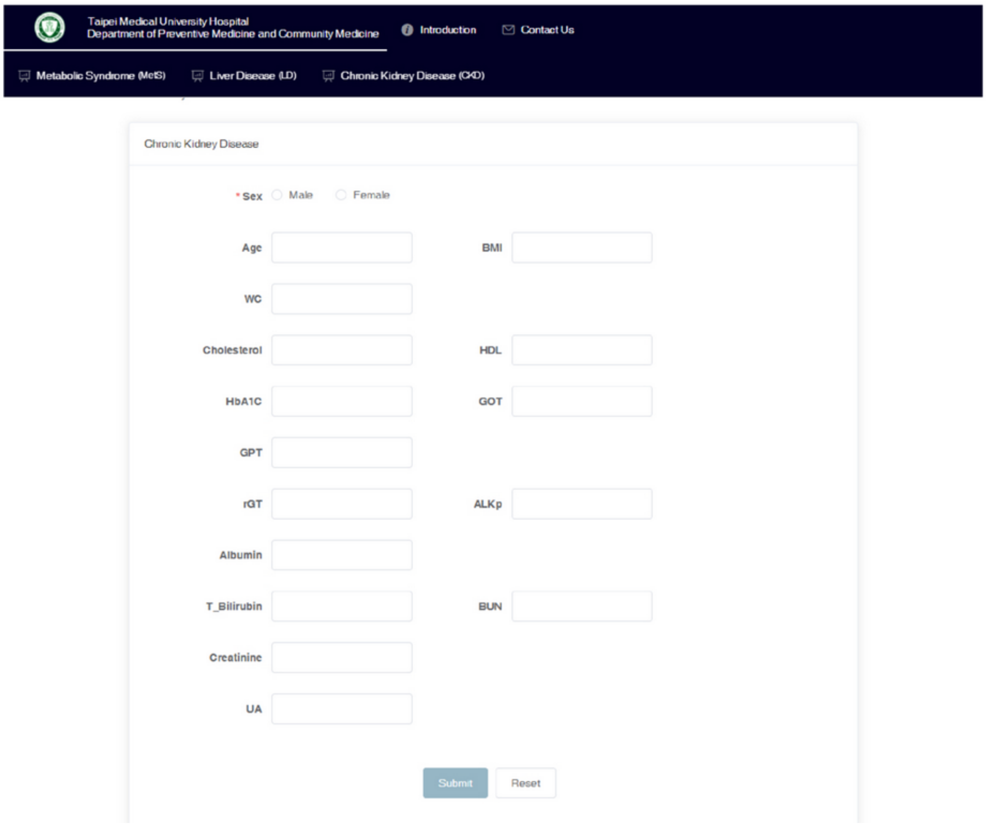
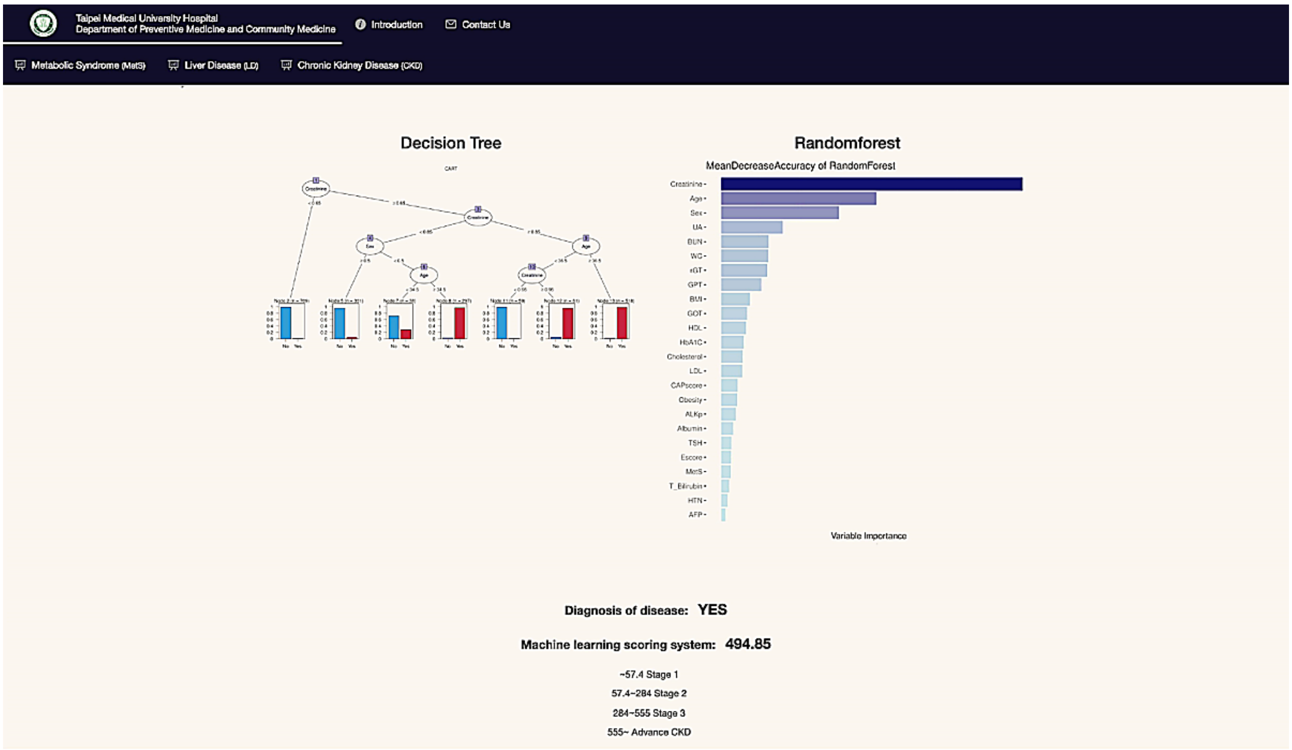


Figure 4. Interface of the input page for disease assessment.



The clinical outcomes established by our database are reported on the website when users have finished entering their medical record data on the website (Figure 5). The CART model and ensemble learning model (random forest) are shown in the output interface. A scoring prediction model obtained using the supervised learning model is also provided online. For unsupervised learning, a color visualization of the clustering heat map depicts a vivid medical pattern of the patient's EMR data, and a record of each user is also constructed using hierarchical clustering with yellow highlights labeling in the heat map (Figure 6). The user will then be classified as more similar to either a healthy subject (green column on the lower left) or an unhealthy subject (orange column on the upper left). A blue bar depicts abnormal values, while a red bar depicts normal values. In addition, on the web system, users can choose to view it as landscape or portrait. The zoom-in and zoom-out functions and the height of the cluster are also dynamic, with users being able to change the settings online to inspect the medical outcomes in detail.

Figure 5. Outcome page for supervised learning models and the scoring system for disease diagnosis.

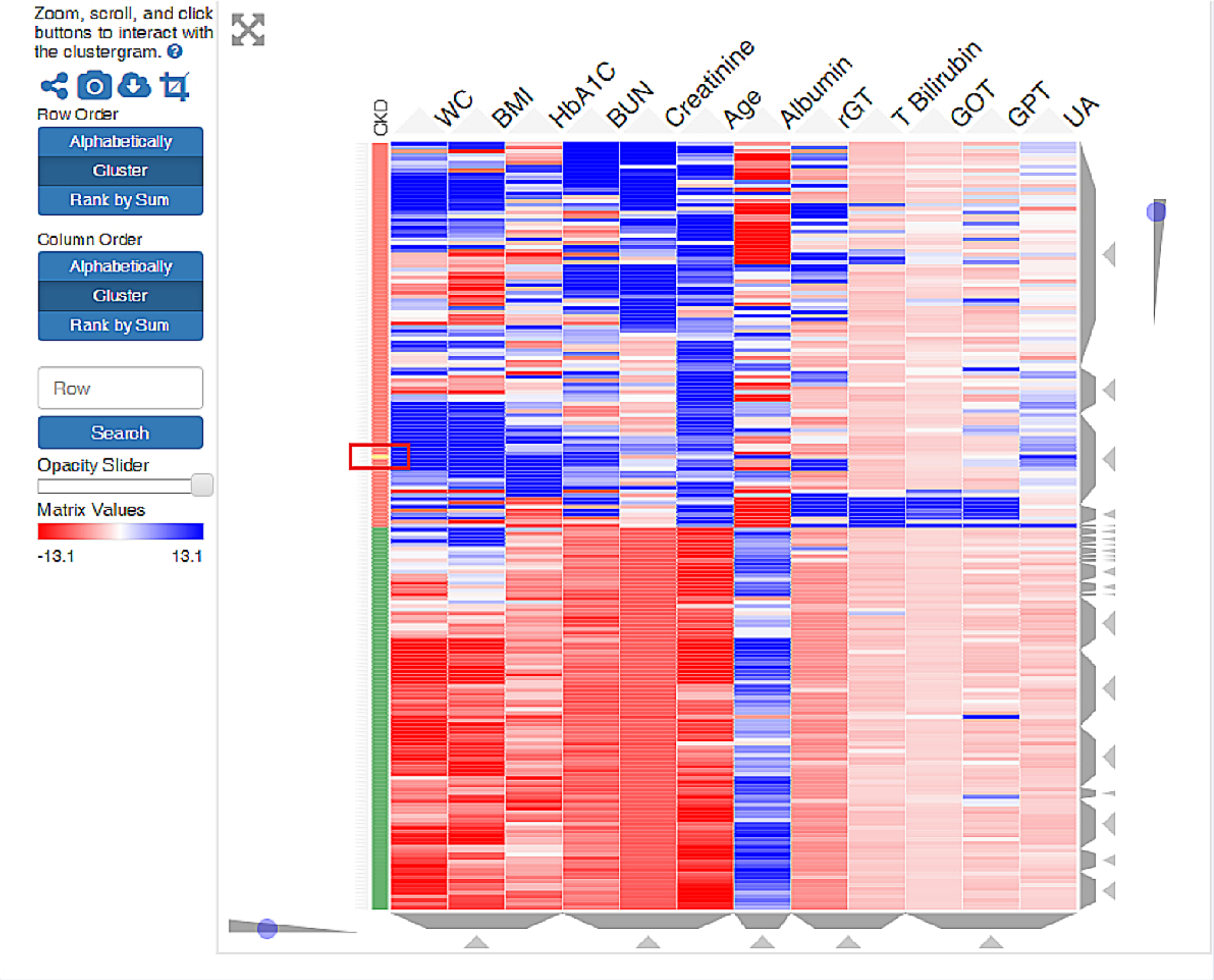


Diagnosis of disease: YES

Machine learning scoring system: 494.85

~57.4 Stage 1
57.4~284 Stage 2
284~555 Stage 3
555~ Advance CKD

Figure 6. Dynamic interactive heat map obtained using unsupervised clustering. Green: healthy patients; orange: CKD patients; blue: normal values; red: abnormal values. The new patients (yellow bar in red rectangle) are compared and clustered into the system's patient database.



Zoom, scroll, and click buttons to interact with the clustergram.

Characteristics of participants in the training and validation data set for MetS can be found in Table 2 and the characteristics of participants of the training set and the validation sets for CKD can be found in Table 3. In general, there are minimal differences in patient characteristics between the training data set and the validation data set for the Taiwanese population of MetS and CKD. However, when comparing the characteristics of the Taiwanese population with other populations (US, Italy, and Japan) for CKD ML performance validation, it was found that there are substantial differences in age and presence of hypertension (Table 3).

Table 2. Characteristics of participants in the training and validation data set for metabolic syndrome.

Characteristics	Training data set (n=904)	Validation data set (n=225)
Sex		
Female, n (%)	411 (45.5)	108 (48.0)
Male, n (%)	493 (54.5)	117 (52.0)
Age, years, median (IQR)	44 (37-50.25)	43 (38-50)
Body mass index, kg/m^2, median (IQR)	23.6 (21.3-25.9)	22.9 (21.2-26)
Waist circumference, cm, median (IQR)	81.75 (74.5-88)	80.5 (74-87)
Albumin, g/dL, median (IQR)	4.6 (4.4-4.8)	4.6 (4.4-4.8)
Alkaline phosphatase, IU/L, median (IQR)	58 (49-69)	58 (48-69)
Glutamic-oxaloacetic transaminase, IU/L, median (IQR)	20 (17-24.25)	20 (17-25)
Glutamate pyruvate transaminase, IU/L, median (IQR)	20 (14-31)	19 (14-28)
Total bilirubin, mg/dL, median (IQR)	0.6 (0.5-0.9)	0.6 (0.4-0.8)
γ-Glutamyl transpeptidase, U/L, median (IQR)	18 (12-27)	17 (11.55-25)
Controlled attenuation parameter (CAP) score, dB/m, median (IQR)	247 (211-284)	241 (216-282)
Elastic modulus (E) score, kPa, median (IQR)	4.2 (3.4-4.9)	4 (3.3-4.8)
Blood urea nitrogen, mg/dL, median (IQR)	12 (10-15)	12 (10-14)
Creatinine, mg/dL, median (IQR)	0.8 (0.6-0.9)	0.8 (0.6-0.9)
Estimated glomerular filtration rate (eGFR) using the Modification of Diet in Renal Disease (MDRD) equation, median (IQR)	90.23 (80.49-104.77)	91.28 (82.72-107.26)
Uric acid, mg/dL, median (IQR)	5.5 (4.5-6.6)	5.4 (4.3-6.6)
Systolic blood pressure, mm Hg, median (IQR)	114 (105-125)	114 (105-126)
Diastolic blood pressure, mm Hg, median (IQR)	73 (67-80)	72 (66-81)
Cholesterol, mg/dL, median (IQR)	189 (165-209)	185 (168-209)
Triglycerides, mg/dL, median (IQR)	90 (65-135.2)	86 (63-126)
High-density lipoprotein, mg/dL, median (IQR)	55(45-67)	54 (47-66)
Low-density lipoprotein, mg/dL, median (IQR)	123 (102-145)	123 (102-145)
Hemoglobin A_{1c}, %, median (IQR)	5.4 (5.2-5.6)	5.4 (5.2-5.5)
Glucose AC, mg/dL, median (IQR)	91 (86-96)	90 (85-95)

Table 3. Characteristics of participants in the training and validation data sets for chronic kidney disease.

Characteristics	Training data set (n=1830)	Validation data set, Taiwan (n=457)	Validation data set, United States (n=4434)	Validation data set, Italy (n=655)	Validation data set, Japan (n=996)
Sex, male, n (%)	902 (49.29)	209 (45.73)	2165 (48.83)	384 (58.63)	696 (69.88)
Chronic kidney disease, n (%)	164 (8.96)	38 (8.32)	410 (9.25)	523 (79.85)	919 (92.27)
Hypertension, n (%)	522 (28.52)	140 (30.63)	1730 (39.02)	599 (91.45)	908 (91.16)
Age, years, median (IQR)	46 (38-55)	45 (37-55)	53 (36-65)	67 (56-74.5)	70 (61-77)
Body mass index, kg/m^2, median (IQR)	23.8 (21.4-26.4)	23.4 (21.3-26.2)	28.6 (24.8-33.5)	28.4 (25.8-31.6)	23.25 (21-25.8)
Waist circumference, cm, median (IQR)	82.5 (75.5-89.5)	81 (75-89)	99.5 (89-111.3)	—[a]	—
Glutamic-oxaloacetic transaminase, IU/L, median (IQR)	21 (17-26)	20 (17-25)	19 (16-24)	—	—
Glutamate pyruvate transaminase, IU/L, median (IQR)	20 (14-30)	19 (13-28)	18 (13-26)	—	—
γ-Glutamyl transpeptidase, U/L, median (IQR)	19 (13-30)	18 (12-33)	21 (15-33)	—	—
Total bilirubin, mg/dL, median (IQR)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0.4 (0.3-0.6)	—	—
Alkaline phosphatase, IU/L, median (IQR)	62 (51-76)	63 (50-78)	75 (62-91)	—	—
Blood urea nitrogen, mg/dL, median (IQR)	13 (11-16)	13 (10-15)	14 (11-18)	28 (21.2-37.3)	—
Creatinine, mg/dL, median (IQR)	0.8 (0.6-1.0)	0.7 (0.6-0.9)	0.85 (0.71-1.01)	1.49 (1.2-1.9)	1.8 (1.2-2.75)
Uric acid, mg/dL, median (IQR)	5.5 (4.5-6.7)	5.4 (4.5-6.5)	5.3 (4.4-6.4)	6.3 (5.2-7.6)	—
Albumin, g/dL, median (IQR)	4.6 (4.4-4.8)	4.6 (4.4-4.8)	4.1 (3.9-4.3)	4 (3.7-4.3)	4 (3.5-4.3)
Cholesterol, mg/dL, median (IQR)	186 (164-210)	185 (160-209)	185 (160-214)	189 (162.5-218)	—
High-density lipoprotein, mg/dL, median (IQR)	52 (44-64)	53 (43-64)	51 (42-61)	—	—
Low-density lipoprotein, mg/dL, median (IQR)	121 (100-145)	120 (100-142)	—	—	—
Hemoglobin A_{1c}, %, median (IQR)	5.4 (5.2-5.6)	5.4 (5.2-5.7)	5.6 (5.3-6)	—	—

[a] Not available.

Table 4 shows the validation performances of supervised learning models in predicting MetS and CKD. In general, it was found that the random forest ML model has higher accuracy than the CART model. Using the random forest ML model, MetS can be predicted with an accuracy of 0.909, and CKD can be predicted up to an accuracy of 0.947. Due to the inconsistent definition of MetS globally, the ML performance of the MetS database has only been validated using the Taiwan data set. However, the ML performances of the CKD database have been validated using data sets from Taiwan, Italy, the United States, and Japan. In general, the CKD database shows good external applicability, and has high AUC for all 4 validation data sets (Taiwan: AUC=0.982; USA: AUC=0.929; Italy: AUC=0.977; Japan: AUC=0.923). However, the validation accuracy and F1 value of CKD prediction differs more substantially, as the unavailable data were excluded from the analysis. When compared to the Taiwanese CKD data set, the respective unavailable data are approximately 6% for the US data set, 50%

for the Italy data set, and 67% for the Japan data set. Therefore, it is observed that the Japanese validation data set has the lowest accuracy (0.743) in predicting CKD, as it also has the highest proportion of unavailable data.

Table 4. The performance of supervised learning models on predicting metabolic syndrome and chronic kidney disease.

Model and disease	Accuracy	Area under the curve (AUC)	F1 score
Classification and regression tree (CART)			
Metabolic syndrome (Taiwan)	0.874	0.887	0.448
Chronic kidney disease (Taiwan)	0.945	0.928	0.965
Random forest			
Metabolic syndrome (Taiwan)	0.909	0.904	0.610
Chronic kidney disease (Taiwan)	0.947	0.982	0.989
Chronic kidney disease (United States)	0.951	0.929	0.679
Chronic kidney disease (Italy)	0.881	0.977	0.920
Chronic kidney disease (Japan)	0.743	0.923	0.838

Discussion

Overview

This ML medical system for three common diseases in family medicine (MetS, CKD, and liver diseases) was constructed from EMR subjects who underwent self-paid health examination. Several ML prediction models are applied to the databases, and the outcomes are summarized and presented visually on the website for users and medical staff. The accuracy of predicting MetS reached 90.9%, and AUC was 0.904 in this system. For chronic kidney disease, the prediction accuracy reached 94.7%, and the AUC was 0.982. In general, users who were invited to test this system rated it with good usability and could easily assess their health online through this web-based ML monitoring system.

CART

Decision trees are an important type of ML algorithm for predictive modeling. They are commonly used in data mining with the objective of creating a model that predicts the dependent variable (the target) based on numerous independent variables [34,37].

A decision tree is a nonparametric ML modeling technique used for regression and classification problems. In classification problems, the target variable is categorical, and the tree is used to identify which group or class a target variable would likely fall into. In regression problems, the target variable is continuous, and the tree is used to predict its value. To find solutions, a decision tree makes a sequential, hierarchical decision about the outcome's variable according to the predictor [55-57].

Hence, CART can provide a visual tree-based diagram for medical practitioners to disseminate health care information to patients. It also helps users to understand the significance of different risk factors for specific diseases. For example, the cut-off controlled attenuation parameter (CAP) score was used to separate patients with MetS and those with other health observations. The CAP score was brought to the attention of users, thereby increasing their awareness of self-health [34].

Random Forest

Random forest, also called random decision forests, is a popular ensemble learning method in ML. Ensemble methods use multiple learning algorithms to improve ML results by combining several decision tree models. This approach allows better predictive performance compared with a single model. Random forest is a parallel ensemble method in which the base learners are generated in parallel. The basic motivation of parallel methods is to exploit independence between the base learners because the error can be reduced dramatically by averaging [58,59]. As random forest provides a bagging technique for feature estimates, it also offers efficient estimates of the test error without incurring the cost of repeated model training associated with cross-validation. Moreover, random forest ranks risk factors in prediction models, which clinicians can use as a reference for diagnosis, and remote users can use to review their risk assessment of related diseases [34,60-62]. For instance, clinicians can refer to significant factors of certain diseases to determine whether those factors exceed the thresholds or not, allowing patients to be more vigilant about their risk of developing such diseases. In addition, sequential ensemble methods such as AdaBoost and XGBoost will be implemented and uploaded to our system in the future.

Clustering

Hierarchical clustering is a widely used unsupervised learning technique that groups data with similar characteristics. Both agglomerative and divisive approaches use dendrograms for the results. A heat map is a color graphical representation of data, which uses a matrix with color gradients to present the similarity of data.

Many studies on genetic bioinformatics and bacterial ecology have used heat maps for the analysis of large and complicated data sets, and some medical studies have used heat maps with clustering to present the relationship between various biomarkers according to their characteristics [34,63-65]. Furthermore, our system provides an interactive clustering heat map for health care. From the perspective of big data, users can evaluate their health status by using ML models and EMR databases. In addition to online health evaluation, in the future, this system could be implemented into different IoMT to assist medical practitioners in achieving real-time health evaluations and monitoring remote patients or patients in specific wards. For the heat map, the EMR data of users were grouped into clusters of patients with diseases in the database; they would then be classified as clinically high-risk objects requiring close attention in the clinical setting [34]. Therefore, whether it is applied in preventive medicine for health management, in a monitoring system for critical care, or in the telemedicine environment, our system can provide real-time monitoring and help predict patient conditions.

Limitations and Future Work

To the best of our knowledge, this is the first web-based machine learning system based on self-paid health examination subjects that can provide an online self-health evaluation for several common diseases (MetS, CKD, and liver diseases). The version 1.0 web-based system still has several limitations that may be improved in the next update. First, the 1.0 system is not yet ready for embedding into a hospital for real-time assessment. We are currently working on an improved system to accept unstructured data input and multimodal data, which are especially essential for the prediction of eye diseases such as macular degeneration. Second, the 1.0 system did not have a user login or account security function. Retrievable prediction and security will be improved as the system is matured for hospital embedment. Third, the 1.0 system does not have whole dynamic analyses such as an interactive decision tree; whole dynamic analyses will be incorporated in subsequent versions to improve communication between the medical staff and patients.

In the future, more clustering algorithms will be implemented in subsequent versions to make the prediction results more robust and reliable. Although the 1.0 system can currently only evaluate three chronic diseases (MetS, CKD, and liver diseases) frequently encountered in family medicine, more chronic disease prediction models, such as those for coronary artery disease, will be added in the near future.

Conclusion

We constructed an ML health monitoring system to offer an online health assessment service to medical units, telemedicine patients, and all health-conscious users worldwide. Our aim is that this system will be implemented in medical centers as a real-time patient monitoring system and provide regular health evaluations for telemedicine patients. Online users can now access our platform and use ML technology to estimate their health status, increasing self-health awareness.

Acknowledgments

This study was supported by Taiwan National Science Foundation Grant NSC108-2314-B-038-073-. No funding bodies had any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire information.
[PDF File (Adobe PDF File), 10106 KB - jmir_v22i6e18585_app1.pdf]

Multimedia Appendix 2
Heat map clustering example and system loading test results.
[DOCX File , 2487 KB - jmir_v22i6e18585_app2.docx]

References

1. Goodman KW. Ethics, Computing, and Medicine: Informatics and the Transformation of Health Care. Cambridge, England: Cambridge University Press; 1997.
2. Kim J. Big Data, Health Informatics, and the Future of Cardiovascular Medicine. J Am Coll Cardiol 2017 Feb 21;69(7):899-902 [FREE Full text] [doi: 10.1016/j.jacc.2017.01.006] [Medline: 28209228]
3. Chien TW, Wang WC, Huang SY, Lai WP, Chow JC. A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. J Med Internet Res 2011 Aug 15;13(3):e61 [FREE Full text] [doi: 10.2196/jmir.1785] [Medline: 21844001]
4. De Beurs DP, de Vries AL, de Groot MH, de Keijser J, Kerkhof AJ. Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. J Med Internet Res 2014;16(9):e207 [FREE Full text] [doi: 10.2196/jmir.3511] [Medline: 25213259]
5. Pan L, Liu G, Mao X, Li H, Zhang J, Liang H, et al. Development of Prediction Models Using Machine Learning Algorithms for Girls with Suspected Central Precocious Puberty: Retrospective Study. JMIR Med Inform 2019 Feb 12;7(1):e11728 [FREE Full text] [doi: 10.2196/11728] [Medline: 30747712]
6. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based Real-Time Case Finding for the Population Health Management of Patients With Diabetes Mellitus: A Prospective Validation of the Natural Language Processing-Based

Algorithm With Statewide Electronic Medical Records. JMIR Med Inform 2016 Nov 11;4(4):e37 [FREE Full text] [doi: 10.2196/medinform.6328] [Medline: 27836816]
7. Jones M, Koziel C, Larsen D, Berry P, Kubatka-Willms E. Progress in the Enhanced Use of Electronic Medical Records: Data From the Ontario Experience. JMIR Med Inform 2017 Feb 22;5(1):e5 [FREE Full text] [doi: 10.2196/medinform.6928] [Medline: 28228372]
8. Installé AJ, Van den Bosch T, De Moor B, Timmerman D. Clinical data miner: an electronic case report form system with integrated data preprocessing and machine-learning libraries supporting clinical diagnostic model research. JMIR Med Inform 2014 Oct 20;2(2):e28 [FREE Full text] [doi: 10.2196/medinform.3251] [Medline: 25600863]
9. Fei Y, Hu J, Gao K, Tu J, Li WQ, Wang W. Predicting risk for portal vein thrombosis in acute pancreatitis patients: A comparison of radical basis function artificial neural network and logistic regression models. J Crit Care 2017 Jun;39:115-123. [doi: 10.1016/j.jcrc.2017.02.032] [Medline: 28246056]
10. Santelices LC, Wang Y, Severyn D, Druzdzel MJ, Kormos RL, Antaki JF. Development of a hybrid decision support model for optimal ventricular assist device weaning. Ann Thorac Surg 2010 Sep;90(3):713-720 [FREE Full text] [doi: 10.1016/j.athoracsur.2010.03.073] [Medline: 20732482]
11. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991 Dec 01;115(11):843-848. [doi: 10.7326/0003-4819-115-11-843] [Medline: 1952470]
12. Artis SG, Mark RG, Moody GB. Detection of atrial fibrillation using artificial neural networks. In: 1991 Proceedings of Computers in Cardiology.: IEEE; 1991 Presented at: Computers in Cardiology (CinC); September 23-26; Venice, Italy p. 173-176 URL: https://ieeexplore.ieee.org/document/169073 [doi: 10.1109/CIC.1991.169073]
13. Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Med Inform Decis Mak 2005 Feb 15;5:3 [FREE Full text] [doi: 10.1186/1472-6947-5-3] [Medline: 15713231]
14. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng Online 2014 Jul 05;13:94 [FREE Full text] [doi: 10.1186/1475-925X-13-94] [Medline: 24998888]
15. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. IEEE J Biomed Health Inform 2017 Dec;21(1):4-21. [doi: 10.1109/JBHI.2016.2636665] [Medline: 28055930]
16. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
17. Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: Opportunities and challenges. Neurocomputing 2017 May;237:350-361. [doi: 10.1016/j.neucom.2017.01.026]
18. Verbrugge LM, Reoma JM, Gruber-Baldini AL. Short-term dynamics of disability and well-being. J Health Soc Behav 1994 Jun;35(2):97-117. [Medline: 8064125]
19. Chatterji S, Byles J, Cutler D, Seeman T, Verdes E. Health, functioning, and disability in older adults--present status and future implications. Lancet 2015 Feb 07;385(9967):563-575 [FREE Full text] [doi: 10.1016/S0140-6736(14)61462-8] [Medline: 25468158]
20. Leavell HR, Clark EG. Textbook of Preventive Medicine. New York: McGraw-Hill; 1958.
21. Jung P, Lushniak BD. Preventive Medicine's Identity Crisis. Am J Prev Med 2017 Mar;52(3):e85-e89. [doi: 10.1016/j.amepre.2016.10.037] [Medline: 28012813]
22. Margolis SA. Preventive healthcare: A core component of Australian general practice. Aust J Gen Pract 2018 Dec;47(12):821 [FREE Full text] [doi: 10.31128/AJGP-11-18-4751] [Medline: 31212397]
23. Hensrud DD. Clinical preventive medicine in primary care: background and practice: 1. Rationale and current preventive practices. Mayo Clin Proc 2000 Feb;75(2):165-172. [doi: 10.4065/75.2.165] [Medline: 10683656]
24. Kassi E, Pervanidou P, Kaltsas G, Chrousos G. Metabolic syndrome: definitions and controversies. BMC Med 2011 May 05;9:48 [FREE Full text] [doi: 10.1186/1741-7015-9-48] [Medline: 21542944]
25. Ding C, Yang Z, Wang S, Sun F, Zhan S. The associations of metabolic syndrome with incident hypertension, type 2 diabetes mellitus and chronic kidney disease: a cohort study. Endocrine 2018 May;60(2):282-291. [doi: 10.1007/s12020-018-1552-1] [Medline: 29492904]
26. Grundy SM. Metabolic syndrome: a multiplex cardiovascular risk factor. J Clin Endocrinol Metab 2007 Feb;92(2):399-404. [doi: 10.1210/jc.2006-0513] [Medline: 17284640]
27. Lorenzo C, Okoloise M, Williams K, Stern MP, Haffner SM, San Antonio Heart Study. The metabolic syndrome as predictor of type 2 diabetes: the San Antonio heart study. Diabetes Care 2003 Nov;26(11):3153-3159. [doi: 10.2337/diacare.26.11.3153] [Medline: 14578254]
28. Lin YJ, Lin CH, Wang ST, Lin SY, Chang SS. Noninvasive and Convenient Screening of Metabolic Syndrome Using the Controlled Attenuation Parameter Technology: An Evaluation Based on Self-Paid Health Examination Participants. J Clin Med 2019 Oct 24;8(11):1775 [FREE Full text] [doi: 10.3390/jcm8111775] [Medline: 31653028]
29. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, National Kidney Foundation. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann Intern Med 2003 Jul 15;139(2):137-147. [doi: 10.7326/0003-4819-139-2-200307150-00013] [Medline: 12859163]

30. Drawz P, Rahman M. Chronic kidney disease. Ann Intern Med 2015 Jun 02;162(11):ITC1-IT16. [doi: 10.7326/AITC201506020] [Medline: 26030647]
31. Webster AC, Nagler EV, Morton RL, Masson P. Chronic Kidney Disease. Lancet 2017 Mar 25;389(10075):1238-1252. [doi: 10.1016/S0140-6736(16)32064-5] [Medline: 27887750]
32. Trettel A, Eissing L, Augustin M. Telemedicine in dermatology: findings and experiences worldwide - a systematic literature review. J Eur Acad Dermatol Venereol 2018 Feb;32(2):215-224. [doi: 10.1111/jdv.14341] [Medline: 28516492]
33. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
34. Yu CS, Lin CH, Lin YJ, Lin SY, Wang ST, Wu JL, et al. Clustering Heatmap for Visualizing and Exploring Complex and High-dimensional Data Related to Chronic Kidney Disease. J Clin Med 2020 Feb 02;9(2):403 [FREE Full text] [doi: 10.3390/jcm9020403] [Medline: 32024311]
35. Breiman L, Friedman JH, Stone CJ, Olshen RA. Classification and Regression Trees. Boca Raton, Florida: CRC Press; Jan 1, 1984.
36. Yu CS, Lin YJ, Lin CH, Wang ST, Lin SY, Lin SH, et al. Predicting Metabolic Syndrome With Machine Learning Models Using a Decision Tree Algorithm: Retrospective Cohort Study. JMIR Med Inform 2020 Mar 23;8(3):e17110 [FREE Full text] [doi: 10.2196/17110] [Medline: 32202504]
37. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data-NHANES 2017-2018. Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention URL: https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017 [accessed 2020-03-27]
38. De Nicola L, Provenzano M, Chiodini P, Borrelli S, Garofalo C, Pacilio M, et al. Independent Role of Underlying Kidney Disease on Renal Prognosis of Patients with Chronic Kidney Disease under Nephrology Care. PLoS One 2015;10(5):e0127071 [FREE Full text] [doi: 10.1371/journal.pone.0127071] [Medline: 25992629]
39. Iimori S, Naito S, Noda Y, Sato H, Nomura N, Sohara E, et al. Prognosis of chronic kidney disease with normal-range proteinuria: The CKD-ROUTE study. PLoS One 2018 Jan 17;13(1):e0190493 [FREE Full text] [doi: 10.1371/journal.pone.0190493] [Medline: 29342207]
40. Peute L, Scheeve T, Jaspers M. Classification and Regression Tree and Computer Adaptive Testing in Cardiac Rehabilitation: Instrument Validation Study. J Med Internet Res 2020 Jan 30;22(1):e12509 [FREE Full text] [doi: 10.2196/12509] [Medline: 32012065]
41. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning. 2006 Presented at: International Conference on Machine Learning; June 25-29; Pittsburgh, Pennsylvania p. 161-168 URL: https://dl.acm.org/doi/10.1145/1143844.1143865 [doi: 10.1145/1143844.1143865]
42. Rokach L, Maimon O. Clustering Methods. In: Data Mining and Knowledge Discovery Handbook. Boston, MA: Springer; 2005:321-352.
43. Ward JH. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 1963 Mar;58(301):236-244. [doi: 10.1080/01621459.1963.10500845]
44. Cormack RM. A Review of Classification. Journal of the Royal Statistical Society. Series A (General) 1971;134(3):321. [doi: 10.2307/2344237]
45. Wilkinson L, Friendly M. The History of the Cluster Heat Map. The American Statistician 2009 May;63(2):179-184. [doi: 10.1198/tas.2009.0033]
46. Perrot A, Bourqui R, Hanusse N, Lalanne F, Auber D. Large interactive visualization of density functions on big data infrastructure. In: 2015 IEEE 5th Symposium on Large Data Analysis and Visualization (LDAV). 2015 Presented at: LDAV 2015 - Big Data Analysis and Visualization; October 25-26; Chicago, IL, USA URL: https://ieeexplore.ieee.org/document/7348077 [doi: 10.1109/ldav.2015.7348077]
47. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R Programming Tools for Plotting Data. 2020. URL: https://cran.r-project.org/web/packages/gplots/index.html [accessed 2020-02-26]
48. Joyia GJ, Liaqat RM, Farooq A, Rehman S. Internet of Medical Things (IOMT): Applications, Benefits and Future Challenges in Healthcare Domain. JCM 2017;12(4):240-247 [FREE Full text] [doi: 10.12720/jcm.12.4.240-247]
49. Istepanian RSH, Sungoor A, Faisal A, Philip N. Internet of m-health Things (m-IoT). In: IET Seminar on Assisted Living 2011. 2011 Presented at: IET Seminar on Assisted Living; April 6; London, UK p. 1-40 URL: https://ieeexplore.ieee.org/document/6183148
50. Boutros-Saikali N, Saikali K, Naoum RA. An IoMT platform to simplify the development of healthcare monitoring applications. In: 2018 Third International Conference on Electrical and Biomedical Engineering, Clean Energy and Green Computing (EBECEGC). 2018 Presented at: EBECEGC 2018: The Third International Conference on Electrical and Biomedical Engineering, Clean Energy and Green Computing (EBECEGC2018); April 25-27; Beirut, Lebanon p. 25-27. [doi: 10.1109/ebecegc.2018.8357124]

51. Lewis JR, Sauro J. The Factor Structure of the System Usability Scale. In: HCD 2009: Human Centered Design. 2009 Presented at: International Conference on Human Centered Design; July 19-24; San Diego, CA, USA p. 94-103 URL: https://link.springer.com/content/pdf/10.1007%2F978-3-642-02806-9.pdf
52. Flavián C, Guinalíu M, Gurrea R. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Information & Management 2006 Jan;43(1):1-14. [doi: 10.1016/j.im.2005.01.002]
53. Yu CS, Chang SS. ML-Doctor. URL: https://www.mldoctor.com.tw [accessed 2020-02-15]
54. T W. Measuring and Interpreting System Usability Scale (SUS). URL: https://uiuxtrend.com/measuring-system-usability-scale-sus/ [accessed 2020-04-01]
55. Friedl MA, Brodley CE. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment 1997 Sep;61(3):399-409. [doi: 10.1016/s0034-4257(97)00049-7]
56. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics 1991;21(3):660-674. [doi: 10.1109/21.97458]
57. Fan CY, Chang PC, Lin JJ, Hsieh JC. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Applied Soft Computing 2011 Jan;11(1):632-644. [doi: 10.1016/j.asoc.2009.12.023]
58. Ho TK. The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Machine Intell 1998;20(8):832-844. [doi: 10.1109/34.709601]
59. Haste T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer; 2009.
60. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2/3:18-22 [FREE Full text]
61. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis 2008 Jan;52(4):2249-2260. [doi: 10.1016/j.csda.2007.08.015]
62. Yang F, Wang HZ, Mi H, Lin CD, Cai WW. Using random forest for reliable classification and cost-sensitive learning for medical diagnosis. BMC Bioinformatics 2009 Jan 30;10 Suppl 1:S22 [FREE Full text] [doi: 10.1186/1471-2105-10-S1-S22] [Medline: 19208122]
63. Benton PH, Ivanisevic J, Rinehart D, Epstein A, Kurczy ME, Boska MD, et al. An Interactive Cluster Heat Map to Visualize and Explore Multidimensional Metabolomic Data. Metabolomics 2015 Aug 01;11(4):1029-1034 [FREE Full text] [doi: 10.1007/s11306-014-0759-2] [Medline: 26195918]
64. Nithya NS, Duraiswamy K, Gomathy P. A Survey on Clustering Techniques in Medical Diagnosis. International Journal of Computer Science Trends and Technology 2013;1(2):17-22 [FREE Full text]
65. Ziemba YC, Lomsadze L, Jacobs Y, Chang TY, Haghi N. Using Heatmaps to Identify Opportunities for Optimization of Test Utilization and Care Delivery. J Pathol Inform 2018;9:31 [FREE Full text] [doi: 10.4103/jpi.jpi_7_18] [Medline: 30294500]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
CAP: controlled attenuation parameter
CART: classification and regression tree
CKD: chronic kidney disease
eGFR: estimated glomerular filtration rate
EMRs: electronic medical records
HMC: Health Management Center
IoMT: Internet of Medical Things
IRB: institutional review board
MDRD: Modification of Diet in Renal Disease
MetS: metabolic syndrome
ML: machine learning
ROC: receiver operating characteristic
TMUH: Taipei Medical University Hospital

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by JY Wu, SC Yeh, Z Yang, B Jin; comments to author 23.03.20; revised version received 13.04.20; accepted 14.05.20; published 05.06.20.

Please cite as:
Yu CS, Lin YJ, Lin CH, Lin SY, Wu JL, Chang SS
Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach
J Med Internet Res 2020;22(6):e18585
URL: https://www.jmir.org/2020/6/e18585
doi:10.2196/18585
PMID:32501272

Original Paper

Patients' Perceptions of Barriers and Facilitators to the Adoption of E-Hospitals: Cross-Sectional Study in Western China

Peiyi Li[1], MSc; Yunmei Luo[1], MSc; Xuexin Yu[2], MSc; Jin Wen[1], PhD; Elizabeth Mason[3]; Weimin Li[4*], PhD; Mohammad S Jalali[3,5*], PhD

[1]Institute of Hospital Management, West China Hospital of Sichuan University, Chengdu, China
[2]Biomedical Big Data Center, West China Hospital of Sichuan University, Chengdu, China
[3]Massachusetts General Hospital's Institute for Technology Assessment, Harvard Medical School, Boston, MA, United States
[4]Department of Respiratory Medicine, West China Hospital of Sichuan University, Chengdu, China
[5]MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, United States
[*]these authors contributed equally

Corresponding Author:
Weimin Li, PhD
Department of Respiratory Medicine
West China Hospital of Sichuan University
Guo Xue Xiang 37
Chengdu, 610041
China
Phone: 86 2885424666
Email: weimi003@scu.edu.cn

Abstract

Background: As an innovative approach to providing web-based health care services from physical hospitals to patients at a distance, e-hospitals (ie, extended care hospitals through the internet) have been extensively developed in China. This closed health care delivery chain was developed by combining e-hospitals with physical hospitals; treatment begins with web-based consultation and registration, and then, patients are diagnosed and treated in a physical hospital. This approach is promising in its ability to improve accessibility, efficiency, and quality of health care. However, there is limited research on end users' acceptance of e-hospitals and the effectiveness of strategies aimed to prompt the adoption of e-hospitals in China.

Objective: This study aimed to provide insights regarding the adoption of e-hospitals by investigating patients' willingness to use e-hospitals and analyzing the barriers and facilitators to the adoption of this technology.

Methods: We used a pretested self-administered questionnaire and performed a cross-sectional analysis in 1032 patients across three hierarchical hospitals in West China from June to August 2019. Patients' sociodemographic characteristics, medical history, current disease status, proficiency with electronic devices, previous experience with web-based health services, willingness to use e-hospitals, and perceived facilitators and barriers were surveyed. Multiple significance tests were employed to examine disparities across four age groups, as well as those between patients who were willing to use e-hospitals and those who were not. Multivariate logistic regression was also performed to identify the potential predictors of willingness to use e-hospitals.

Results: Overall, it was found that 65.6% (677/1032) of participants were willing to use e-hospitals. The significant predictors of willingness to use e-hospitals were employment status (P=.02), living with children (P<.001), education level (P=.046), information technology skills (P<.001), and prior experience with web-based health care services (P<.001), whereas age, income, medical insurance, and familiarity with e-hospitals were not predictors. Additionally, the prominent facilitators of e-hospitals were convenience (641/677, 94.7%) and accessibility to skilled medical experts (489/677, 72.2%). The most frequently perceived barrier varied among age groups; seniors most often reported their inability to operate technological devices as a barrier (144/166, 86.7%), whereas young participants most often reported that they avoided e-hospital services because they were accustomed to face-to-face consultation (39/52, 75%).

Conclusions: We identified the variables, facilitators, and barriers that play essential roles in the adoption of e-hospitals. Based on our findings, we suggest that efforts to increase the adoption of e-hospitals should focus on making target populations accustomed to web-based health care services while maximizing ease of use and providing assistance for technological inquiries.

(J Med Internet Res 2020;22(6):e17221) doi:10.2196/17221

KEYWORDS

innovation adoption; e-hospital; internet hospital; eHealth; barriers; facilitators

Introduction

The vast majority of high-quality medical resources (eg, well-trained medical workers and advanced medical equipment) are focused in tertiary hospitals in the urban cities of China. This has led to a lack of patients in secondary hospitals and primary health care centers (PHCs), as well as overcrowding in tertiary hospitals. Consequently, patients in tertiary hospitals have long registration and queue times, long waiting times, long dispensary and payment queue times, and short physician visit times, coined as the “three long and one short” condition [1].

In order to reduce the unbalanced distribution of medical resources, China has developed a hierarchical medical treatment system, in which medical institutions of various levels receive and treat patients according to the severity and urgency of their diseases [2,3]. Specifically, tertiary hospitals treat patients with complex and urgent conditions, whereas secondary hospitals treat patients with common diseases. Providers in PHCs have the responsibility of chronic disease management, and they refer patients to specialists or hospitals when necessary [4]. However, despite a great amount of financial investment [5], many PHCs still have poor clinical performance and deficient medical knowledge, as they lack well-trained and qualified physicians [6,7]. Specifically, previous research has found that more than half of the health practitioners in China do not have a bachelor’s degree, and the education level is particularly low for providers in rural areas [8]. Therefore, crowds of patients continue to travel to overfilled top-level hospitals in pursuit of quality care at the cost of escalating health care expenditure and time [9].

Major developments in information and communication technology (ICT) and increased prevalence of electronic devices have enabled innovations in health delivery in developed countries, such as the United States, the United Kingdom, and Canada, particularly for low-resource and underserved communities [10-13]. The World Health Organization defines electronic health (eHealth) as “the use of ICT to support the delivery of health services and the management of health systems” [14]. Studies on eHealth interventions have found positive effects for various diseases, including symptom reduction, improved health care accessibility, and higher patient and clinician satisfaction [15-18].

Similar benefits have resulted from eHealth efforts in China as well [19]. As a priority eHealth project, e-hospitals were proposed by the National Health Commission of China as an innovative approach to health care service delivery in 2015 [20], and they were expected to alleviate the dilemmas regarding accessibility, cost, and quality [21]. There are two major kinds of e-hospitals. There are e-hospitals that are administrated by physical tertiary hospitals, where patients are able to reach physicians in these tertiary hospitals via the internet and are referred to PHCs or secondary hospitals in the region. Other e-hospitals are established by investment companies and have registered physicians from all over the country [22].

E-hospitals, also known as “extended care hospitals,” take the form of a smartphone app or website, and they represent a new approach to outpatient service delivery through the internet [23]. E-hospitals have strengthened the communication pathway between health specialists and patients by overcoming time and distance barriers. Through this technology, health care professionals are able to provide high-quality continuous services to their patients, improve disease management, have more frequent follow-ups, answer patient concerns, and receive appointment requests.

To access these health care services, individuals need to register their information, such as name, age, and gender, at their e-hospital of choice. Once registered, users are able to describe their condition via written text, voice message, or video, and upload relevant documents and laboratory images. This information is conveyed to specialized physicians in tertiary hospitals through a chat platform where the patient and physician can have a remote conversation. Depending on the condition, the physician may issue an e-prescription for users to purchase medicines at the linked web-based pharmacy or make an appointment at a nearby PHC or secondary hospital and provide instructions for treatment remotely. Lastly, e-hospitals can provide posttreatment care and monitoring for patients after they are discharged from a physical hospital [24]. E-hospitals make it possible for patients, regardless of their location, to communicate with skilled medical experts, which thereby improves the efficiency and accessibility of quality medical services while saving indirect health costs [23].

Over the past 4 years, 294 e-hospitals have been established in China to help distribute the health services of tertiary hospitals to patients in distant underserved areas [25]. However, this new medical technology has had a lower adoption rate than expected, as it did not appear to rid tertiary hospitals of the inefficiencies described by the “three long and one short” condition [26]. Furthermore, a large number of these e-hospital programs ended in the early stages due to lack of users [27,28].

The implementation cycle plays a critical role in the effective use of this new technology [29]. A primary issue regarding the implementation cycle of e-hospitals is the inadequate understanding of how to best facilitate individuals’ and organizations’ adoption of the technology [30]. Organizations may fail to successfully implement e-hospitals as a result of a lack of readiness [31,32], which is defined as the preparedness of health care users to adopt eHealth for the provision and management of health services [33]. A number of factors may impact user readiness, which thereby affects the success of eHealth programs and whether the desired health outcomes are achieved [34,35]. At the organizational level, factors, such as coordination with concerned stakeholders and proper training, can impact readiness for eHealth implementation [36,37].

Critical readiness-determining factors for health care providers include the intensity of their current workload and the perceived quality of the new technology [38]. Unsurprisingly, patient readiness is also essential to the success of the implementation cycle of eHealth interventions. Relevant research on this topic has been conducted in Western countries, and actions that overcome financial and technical barriers to facilitate successful adoption should be undertaken [39-42]. Even though these patient attitudes toward eHealth projects can to some extent be contextualized to other countries [43], the lack of studies directly assessing the Chinese population leads to ambiguous policy suggestions for health care administrators and managers of e-hospitals in China.

Taking all these factors into account, this study aimed to (1) examine how familiar patients in Western China are with e-hospitals and how willing they are to adopt the technology; (2) investigate the potential factors that influence patient adoption of e-hospitals; and (3) explore various patient perceptions of the advantages and disadvantages of using e-hospitals. We intended for our results to provide evidence-based insights for policy regarding the integration of e-hospitals into China's health care system and ultimately deepen the adoption of e-hospitals in China.

Methods

Setting and Ethical Consent

This cross-sectional study was conducted in three hospitals in Chengdu, the capital of Sichuan province, located in West China, with a population of roughly 83.4 million people [44]. For context, there are more PHCs in Sichuan than any other province in China [45]. In an attempt to mitigate selection bias, the study included patients from all hospital tiers as follows: a tertiary hospital (West China Hospital of Sichuan University), a secondary hospital (First People's Hospital of Longquan District), and PHCs (Community Health Centers of Chenghua District). Institutional review board approval was obtained from the Research Ethics Committee of the West China Hospital, Sichuan University.

Study Design and Population

All adult patients (aged 18 years or above) presenting at the aforementioned hospitals from June to September 2019 were asked to complete an anonymous self-administered questionnaire in the physicians' offices of corresponding departments. No restriction for patients' diseases or health care status was an approach to remediate selection bias in our study. Respondents were provided with information about the objectives and scope of the survey and were asked, without incentives, to provide consent to participate in the study. Those who agreed to participate were administered the questionnaire by a research investigator, and careful attention was paid that the participants were not in physicians' offices to avoid potential discomfort. Since acute illnesses are explicitly excluded from e-hospital treatment, patients who went through an emergency procedure were not eligible to participate.

Instrumentation

Our survey (Multimedia Appendix 1) involved multiple-choice questions divided into the following five categories: (1) sociodemographic and disease characteristics; (2) current usage of electronic devices; (3) previous experience with web-based health care services; (4) willingness to use e-hospitals; and (5) perceived facilitators or barriers regarding e-hospital use. More specifically, depending on the response to part 4, in part 5, participants were asked either what encourages their e-hospital usage or what deters it.

We developed the questionnaire based on a literature review of relevant studies [46,47]. To ensure its validity, we pilot tested the questionnaire using two complementary approaches. First, the questionnaire was evaluated by five experts in the fields of hospital administration, medical informatics, and health care policy, and suggestions were used to increase clarity. Second, we pilot tested the questionnaire with 20 patients and used their feedback to make adjustments to the questionnaire.

Measures

Our survey consisted of multiple-choice questions regarding sociodemographic, medical, and eHealth usage–related variables. The variables analyzed were age, gender, level of education, employment status (working vs retired), household location (Chengdu vs outside Chengdu), monthly income, and type of medical insurance. Age ranges were selected based on the categorization by the National Bureau of Statistics of China. To clarify, urban employee health insurance is mainly funded by a person's employer, whereas rural and urban resident health insurances are mainly funded by government subsidies [48]. The response options for monthly income were stratified into four groups (<2000; 2000-5999; 6000-10,000; and >10,000 CNY) [49]. Furthermore, data about whether participants had chronic diseases were recorded to represent the long-term health demand for chronic disease management. Whether a participant had an operation was also recorded to indicate the eventual need for postoperative rehabilitation care.

Furthermore, information technology (IT) skills and living with children (yes vs no) were surveyed to measure participants' technical skills and the potential for technical support from younger generations, respectively. Overall IT proficiency was measured by evaluating responses regarding "number of owned electronic devices," "capability to connect to Wi-Fi," and "capability to install apps." Each category was given a score of 0 to 5, and then, the scores were summed in order to generate an overall IT score. For binary questions, a score of 0 was given to a "no" response and 5 was given to a "yes" response.

Additionally, participants were asked if they had previous experience with web-based health services (yes vs no). Respondents with previous experience were then asked to indicate the extent of their satisfaction on a 5-point Likert scale (5, extremely satisfied; 4, satisfied; 3, neutral; 2, dissatisfied; and 1, extremely dissatisfied). For evaluation of familiarity with e-hospitals, another 5-point Likert scale was used (5, extremely familiar; 4, quite familiar; 3, know a little bit; 2, only heard the term; and 1, never heard the term).

Data Collection and Entry

Four trained research staff distributed the paper-based questionnaires to participants. They followed a verbal script and were instructed to address any potential doubts related to the topics covered by the questionnaire. To confirm that the questionnaire was completely filled out, researchers reviewed the responses immediately after the participant completed the questionnaire.

Two authors (PL and YL) experienced in data entry independently entered all case record data into EpiData (version 3.1, EpiData Software). The two Excel spreadsheets were then compared and discrepancies were resolved by checking the original questionnaires, eventually reaching a consensus between the two authors.

Data Analysis

For statistical analysis, categorical variables were expressed as frequencies and percentages, and continuous variables were expressed as mean values with standard deviations. Descriptive analyses, including the chi-square test, Kruskal-Wallis test, Fisher test, *t* test, and variance analysis, were performed according to the data characteristics. Specifically, multiple significance tests were employed to examine whether there existed differences across age groups. Similarly, statistical significance tests were conducted to examine the association between all other variables and the willingness to use e-hospitals.

Furthermore, a multivariate logistic regression analysis with a range of variables was performed to identify potential indicators of patients' willingness to use e-hospitals. All variables in the descriptive analysis of the willingness to use e-hospitals were included in the multivariate logistic regression model. Variables with a two-tailed *P value* <.05 were considered statistically significant. Additionally, a chi-square test was employed to analyze age-related variations in perceived facilitators and barriers.

All statistical analyses were performed using SPSS (version 25, IBM Corp).

Results

Participants

A total of 1108 patients completed the survey after 43 patients refused to respond. Seventy-six patients were excluded due to incomplete responses. Overall, this study had a high response rate of 89.7% (1032 patients).

Sociodemographic and Medical History

The descriptive analysis in Table 1 shows that socioeconomic attributes varied across age groups. Among the participants, 44.4% (458/1032) were male. Among those aged 18 to 34 years, the number of female participants (166/260, 63.8%) was nearly double that of male participants (94/260, 36.2%), although there appeared to be slightly more male participants (139/276, 50.4%) than female participants (137/276, 49.6%) among those aged 65 years or older. In addition, younger participants were more educated than older participants (*P*<.001). Specifically, 64.2% (167/260) of participants aged 18 to 34 years had attended college or above, whereas this was only true for 14.1% to 37.0% of participants in the other three age groups.

The proportion of retired individuals was significantly higher among those aged 65 or older (270/276, 97.8%) than among the other age groups (50-64 years: 164/258, 63.6%; 35-49 years: 21/238, 8.8%; 18-34 years: 4/260, 1.5%). In addition, participants' monthly income appeared to decline with age (*P*<.001).

Moreover, 55.7% (575/1032) of participants reported living outside of Chengdu. Participants aged 35 years or older appeared to be more likely than younger participants to seek health care services in Chengdu despite not living there (18-34 years: 109/260, 41.9%; 35-49 years: 150/238, 63.0%; 50-64 years: 160/258, 62.0%; 65 years or older: 156/276, 56.5%; *P*<.001). Additionally, nearly half of the older participants did not live with children (50-54 years: 142/256, 55.0%; 65 years or older: 128/276, 46.4%; *P*<.001).

Table 1 also shows that medical history varied across age groups. Older participants appeared to be more likely to have chronic diseases; over 64.9% (179/276) of those aged 65 years or older reported having chronic diseases, whereas this was reported by only 8.5% (22/260) of those aged 18 to 34 years (*P*<.001). Nearly half of the participants (500/1032, 48.4%) underwent an inpatient surgery, and of these, the majority were aged 35 to 64 years. Finally, younger participants more frequently had employee and urban resident insurance than rural and other types of insurance, whereas older participants more frequently had rural resident insurance (*P*<.001).

Current Usage of Electronic Devices and Web-Based Health Care Services

Results in Table 2 show that younger participants were generally more active in web-based activities. Overall, the mean number of electronic devices owned by those aged 18 to 34 years was 2.4 (SD 1.24) and those aged 65 years or older was 0.54 (SD 0.75) (*P*<.001). Meanwhile, 81.9% (226/276) of those aged 65 years or older reported that they were unable to connect their electronic devices to Wi-Fi and 92.4% (255/276) of these participants also indicated that they did not know how to install a new app.

In addition, there was a statistically significant association between previous usage of web-based medical services and age (*P*<.001). Specifically, 90.8% (236/262) of participants aged 18 to 34 years reported that they had at least once received health services over the internet, whereas this was reported by only 15.9% (44/276) of participants aged 65 years or older. Meanwhile, 83.1% (463/557) of current users indicated that they were satisfied with their web-based medical experiences. Lastly, the proportion of individuals who had never heard of e-hospitals was significantly lower among those aged 18 to 34 years (59/260, 22.7%) than among those aged 65 years or older (241/276, 87.3%).

Table 1. Sociodemographic and medical history of the study participants.

Characteristic	Total value	Age stratification in years				*P* value
		18-34	35-49	50-64	≥65	
Sample size, n	1032	260	238	258	276	
Gender, n (%)						.002[a]
Male	458 (44.4)	94 (36.2)	98 (41.2)	127 (49.2)	139 (50.4)	
Female	574 (55.6)	166 (63.8)	140 (58.8)	131 (50.8)	137 (49.6)	
Education level, n (%)						<.001[b]
Primary school or below	277 (26.8)	6 (2.3)	40 (16.8)	96 (37.2)	135 (48.9)	
Junior high school	218 (21.1)	32 (12.3)	62 (26.1)	73 (28.3)	51 (18.5)	
Senior high school	201 (19.5)	55 (21.2)	48 (20.2)	47 (18.2)	51 (18.5)	
College or above	336 (32.6)	167 (64.2)	88 (36.9)	42 (16.3)	39 (14.1)	
Employment status, n (%)						<.001[c]
Working	573 (55.5)	256 (98.5)	217 (91.2)	94 (36.4)	6 (2.2)	
Retired	459 (44.5)	4 (1.5)	21 (8.8)	164 (63.6)	270 (97.8)	
Monthly income (CNY), n (%)						<.001[b]
<2000	381 (36.9)	43 (16.5)	60 (25.2)	139 (53.9)	139 (50.4)	
2000-5999	418 (40.5)	126 (48.5)	108 (45.4)	89 (34.5)	95 (34.4)	
6000-9999	158 (15.3)	67 (25.8)	41 (17.2)	20 (7.8)	30 (10.9)	
≥10,000	75 (7.3)	24 (9.2)	29 (12.2)	10 (3.9)	12 (4.3)	
Home location, n (%)						<.001[a]
Chengdu	457 (44.3)	151 (58.1)	88 (37.0)	98 (38.0)	120 (43.5)	
Outside of Chengdu	575 (55.7)	109 (41.9)	150 (63.0)	160 (62.0)	156 (56.5)	
Living with children, n (%)						<.001[a]
Yes	728 (70.5)	250 (96.2)	214 (89.9)	116 (45.0)	148 (53.6)	
No	304 (29.5)	10 (3.8)	24 (10.1)	142 (55.0)	128 (46.4)	
Having a chronic disease, n (%)						<.001[a]
Yes	381 (36.9)	22 (8.5)	57 (23.9)	123 (47.7)	179 (64.9)	
No	651 (63.1)	238 (91.5)	181 (76.1)	135 (52.3)	97 (35.1)	
Undergone surgery, n (%)						<.001[a]
Yes	500 (48.4)	105 (40.4)	135 (56.7)	148 (57.4)	112 (40.6)	
No	532 (51.6)	155 (59.6)	103 (43.3)	110 (42.6)	164 (59.4)	
Medical insurance, n (%)						<.001[c]
Employee insurance	456 (44.2)	142 (54.6)	113 (47.5)	95 (36.8)	106 (38.4)	
Urban resident insurance	163 (15.7)	51 (19.6)	35 (14.7)	49 (19.0)	28 (10.1)	
Rural resident insurance	368 (35.7)	48 (18.5)	80 (33.6)	109 (42.3)	131 (47.5)	
Others	45 (4.4)	19 (7.3)	10 (4.2)	5 (1.9)	11 (4.0)	

[a] Chi-square test.

[b] Kruskal-Wallis test.

[c] Fisher test.

Table 2. Current usage of electronic equipment and web-based health care services among study participants.

Characteristic	Total value	Age stratification in years				*P* value
		18-34	35-49	50-64	≥65	
Sample size–all, n	1032	260	238	258	276	
Number of electronic devices, mean (SD)	1.51 (4.6)	2.4 (1.24)	1.9 (1.22)	1.28 (1.07)	0.54 (0.75)	<.001[a]
Able to connect to Wi-Fi, n (%)						<.001[b]
Yes	671 (65.0)	259 (99.6)	218 (91.6)	144 (55.8)	50 (18.1)	
No	361 (35.0)	1 (0.4)	20 (8.4)	114 (44.2)	226 (81.9)	
Able to install apps, n (%)						<.001[b]
Yes	558 (54.1)	257 (98.8)	186 (78.2)	94 (36.4)	21 (7.6)	
No	474 (45.9)	3 (1.2)	52 (21.8)	164 (63.6)	225 (92.4)	
Information technology skills[c] score, mean (SD)	7.5 (5.5)	12.3 (1.5)	10.4 (3.7)	6.0 (5.2)	1.8 (3.4)	<.001[a]
Experience of web-based medical services, n (%)						<.001[d]
Yes	557 (54.0)	236 (90.8)	171 (71.8)	106 (41.1)	44 (15.9)	
No	475 (46.0)	24 (9.2)	67 (28.2)	152 (58.9)	232 (84.1)	
Degree of knowledge about e-hospitals						<.001[b]
Very familiar with	37 (3.6)	22 (8.5)	11 (4.6)	4 (1.6)	0 (0)	
Know a better bit	46 (4.5)	25 (9.6)	16 (6.7)	3 (1.2)	2 (0.7)	
Know a good bit	146 (14.1)	73 (28.1)	48 (20.2)	17 (6.6)	8 (2.9)	
Only heard of	206 (20.0)	81 (31.2)	55 (23.1)	45 (17.4)	25 (9.1)	
Never heard of	597 (57.8)	59 (22.7)	108 (45.4)	189 (73.3)	241 (87.3)	
Sample size–web-based medicine users, n	557	236	171	106	44	
Satisfaction with the web-based medical experience, n (%)						.01[b]
Extremely satisfied	199 (35.7)	80 (33.9)	63 (36.8)	47 (44.3)	9 (20.5)	
Satisfied	262 (47.1)	116 (49.2)	80 (46.8)	45 (42.5)	21 (47.7)	
Neutral	80 (14.3)	35 (14.8)	26 (15.2)	10 (9.4)	9 (20.5)	
Dissatisfied	10 (1.8)	3 (1.3)	1 (0.6)	2 (1.9)	4 (9.0)	
Extremely dissatisfied	6 (1.1)	2 (0.8)	1 (0.6)	2 (1.9)	1 (2.3)	

[a] Analysis of variance.

[b] Fisher test.

[c] "Information technology skills" was a combined result of the first, second, and third questions in the relevant part.

[d] Chi-square test.

Willingness to Use E-Hospitals

It was found that 65.6% (677/1032) of participants were willing to use e-hospitals to manage their disease (Table 3). The results suggested that willingness to use e-hospitals was associated with age (*P*=.04), education level (*P*<.001), employment status (*P*<.001), monthly income (*P*<.001), living with children (*P*<.001), medical insurance type (*P*<.001), chronic diseases (*P*<.001), skillful IT operation (*P*<.001), previous experience of web-based health care services (*P*<.001), and familiarity with e-hospitals (*P*<.001) (Table 3).

Table 3. Willingness of participants to use e-hospitals.

Characteristic	Total value	Willingness to use e-hospitals		*P* value
		Yes	No	
Sample size, n (%)	1032	677	355	
Age (years), mean (SD)	50.83 (18.1)	46.49 (16.5)	59.12 (18.1)	.04[a]
Age stratification (years), n (%)				<.001[b]
18-34	260 (25.2)	208 (80.0)	52 (20.0)	
35-50	238 (23.1)	192 (80.7)	46 (19.3)	
50-64	258 (25.0)	167 (64.7)	91 (35.3)	
≥65	276 (26.7)	110 (39.9)	166 (60.1)	
Gender, n (%)				.84[b]
Male	458 (44.4)	302 (65.9)	156 (34.1)	
Female	574 (55.6)	375 (65.3)	199 (34.7)	
Education level, n (%)				<.001[b]
Primary school or below	277 (26.8)	102 (36.8)	175 (63.2)	
Junior high school	218 (21.1)	151 (69.3)	67 (30.7)	
Senior high school	201 (19.5)	148 (73.6)	53 (26.4)	
College or above	336 (32.6)	276 (82.1)	60 (17.9)	
Employment status, n (%)				<.001[b]
Working	573 (55.5)	444 (77.5)	129 (22.5)	
Retired	459 (44.5)	233 (50.8)	226 (49.2)	
Monthly income (CNY), n (%)				<.001[b]
<2000	381 (36.9)	189 (49.6)	192 (50.4)	
2000-5999	418 (40.5)	305 (73.0)	113 (27.0)	
6000-9999	158 (15.3)	121 (76.6)	37 (23.4)	
≥10,000	75 (7.3)	62 (82.7)	13 (17.3)	
Home location, n (%)				.18[b]
Chengdu	457 (44.3)	310 (67.8)	147 (32.2)	
Outside of Chengdu	575 (55.7)	367 (63.8)	208 (36.2)	
Living with children, n (%)				<.001[b]
Yes	728 (70.5)	537 (73.8)	191 (26.2)	
No	304 (29.5)	140 (46.1)	164 (53.9)	
Medical insurance, n (%)				<.001[b]
Employee medical insurance	456 (4.2)	347 (76.1)	109 (23.9)	
Urban resident medical insurance	163 (15.7)	116 (71.2)	47 (28.8)	
Rural cooperative medical insurance	368 (35.7)	181 (49.2)	187 (50.8)	
Others	45 (4.4)	33 (73.3)	12 (26.7)	
Having a chronic disease, n (%)				<.001[b]
Yes	651 (63.1)	466 (71.6)	185 (28.4)	
No	381 (36.9)	211 (55.4)	170 (44.6)	
Undergone surgery, n (%)				.36[b]
Yes	500 (48.4)	335 (67.0)	165 (33.0)	

Characteristic	Total value	Willingness to use e-hospitals		*P* value
		Yes	No	
No	532 (51.6)	342 (64.3)	190 (35.7)	
Information technology skills score, mean (SD)	6.0 (4.6)	9.3 (4.8)	4.0 (5.1)	.002[a]
Experience of web-based medical services, n (%)				<.001[b]
Yes	557 (54.0)	474 (85.1)	83 (14.9)	
No	475 (46.0)	203 (42.7)	272 (57.3)	
Degree of knowledge about e-hospitals, n (%)				<.001[c]
Very familiar with	37 (3.6)	36 (97.3)	1 (2.7)	
Know a better bit	46 (4.5)	40 (87.0)	6 (13.0)	
Know a good bit	146 (14.1)	118 (80.8)	28 (19.2)	
Only heard of	206 (20.0)	167 (81.1)	39 (18.9)	
Never heard of	597 (57.8)	316 (52.9)	281 (47.1)	

[a] *t* test.

[b] Chi-square test.

[c] Fisher test.

Table 4 presents the results of the multivariate logistic regression analysis of the willingness to use e-hospitals (all variables from Table 3 were included). The results showed that age did not make a difference after controlling for covariates in the model. Similarly, variations in monthly income, medical insurance type, chronic diseases (yes vs no), and familiarity with e-hospitals could be explained by other covariates.

In contrast, employment status, education level, living with children, IT skills, and previous experience with web-based health care were closely associated with the willingness to use e-hospitals (Table 4). Specifically, employed participants were 1.88 times more likely to be willing to use e-hospitals compared to retired participants (95% CI 1.11-3.18) after adjusting for all other covariates in the model. Furthermore, participants with higher education were more likely to be willing to use e-hospitals (junior high school: OR 1.95, 95% CI 1.21-3.15; senior high school: OR 1.84, 95% CI 1.00-3.37; college or above: OR 2.16, 95% CI 1.09-4.28). In addition, participants living with children had a higher likelihood of being willing to use e-hospitals compared to those living without children (OR 1.88, 95% CI 1.34-2.64). Furthermore, participants with higher IT skills had a higher probability of being willing to use e-hospitals (OR 1.11, 95% CI 1.05-1.17). Lastly, participants with previous experience of using web-based medical services were 2.69 times more likely to be willing to use e-hospitals (OR 2.77, 95% CI 1.80-4.26).

Table 4. Multivariate logistic regression of the willingness to use e-hospitals.

Independent variable[a]	Coefficient	Wald χ^2	df	OR (95% CI)	*P* value
Constant	0.796	0.419	1	2.22	N/A[b]
Age	−0.003	0.145	1	0.10 (0.98-1.01)	.70
Gender (male vs female)	−0.251	2.417	1	0.78 (0.57-1.07)	.12
Education level					
Primary school or below	N/A	N/A	N/A	Reference	N/A
Junior high school	0.667	7.407	1	1.95 (1.21-3.15)	.006
Senior high school	0.609	3.878	1	1.84 (1.00-3.37)	.049
College or above	0.768	4.822	1	2.16 (1.09-4.28)	.03
Employment status (working vs retired)	0.632	5.545	1	1.88 (1.11-3.18)	.02
Monthly income (CNY)					
<2000	N/A	N/A	N/A	Reference	N/A
2000-5999	−0.046	0.038	1	0.96 (0.60-1.51)	.85
6000-9999	−0.177	0.298	1	0.84 (0.44-1.58)	.59
≥10,000	0.039	0.008	1	1.04 (0.43-2.49)	.93
Living with children (yes vs no)	0.632	13.272	1	1.88 (1.34-2.64)	.001
Home location (Chengdu vs outside Chengdu)	−0.318	3.026	1	0.73 (0.51-1.04)	.08
Medical insurance					
Employee medical insurance	N/A	N/A	N/A	Reference	N/A
Urban resident medical insurance	0.026	0.011	1	1.03 (0.63-1.67)	.92
Rural cooperative medical insurance	−0.056	0.047	1	0.95 (0.57-1.57)	.83
Others	0.024	0.003	1	1.02 (0.73-1.46)	.95
Having a chronic disease (yes vs no)	0.032	0.032	1	1.03 (0.73-1.46)	.86
Undergone surgery (yes vs no)	0.153	0.823	1	1.17 (0.84-1.62)	.36
Information technology skills score	0.102	13.843	1	1.11 (1.05-1.17)	<.001
Experience with web-based medical services (yes vs no)	1.017	21.282	1	2.77 (1.80-4.26)	<.001
Degree of knowledge about e-hospitals					
Very familiar with	N/A	N/A	N/A	Reference	N/A
Know a better bit	−1.680	2.286	1	0.19 (0.02-1.66)	.13
Know a good bit	−1.960	3.511	1	0.14 (0.02-1.09)	.06
Only heard of	−1.752	2.831	1	0.17 (0.02-1.34)	.09
Never heard of	−1.916	3.385	1	0.15 (0.02-1.13)	.07

[a] Specific analysis for the multivariate logistic regression model: 2lnL=1036.027; Hosmer and Lemeshow test: χ^2=7.029, *P*=.53.

[b] N/A: not applicable.

Perceived Facilitators for Users and Barriers for Nonusers

The vast majority of participants considered convenience (641/677, 94.7%) to be a major facilitator for e-hospital adoption. The next most popularly agreed upon facilitator was improved access to skilled experts (489/677, 72.2%), followed by improved health outcomes (184/677, 27.2%), privacy protection (180/677, 26.6%), and active participation in disease self-management (144/677, 21.3%). Furthermore, differences in terms of perceived facilitators and barriers appeared to exist across age groups (Figure 1). It was notable that younger participants (aged 18-34 years) showed greater interest in improved health outcomes (79/208, 37.98%; *P*<.001), privacy protection (103/208, 49.5%; *P*<.001), and disease self-management (74/208, 35.8%; *P*<.001) as compared with the other three age groups. In addition, participants aged 50 to 64 years appeared to show greater interest in convenience

(165/167, 98.8%; *P*=.04). There was no relevant difference in interest regarding improved accessibility to skilled experts across the age groups (*P*=.32).

The barriers reported were inability to operate electronic devices (238/355, 67.0%), familiarity with face-to-face health care (108/355, 30.4%), doubts regarding the authenticity and reliability of e-hospitals (86/355, 24.2%), useless perceptions about e-hospitals (37/355, 10.4%), and concerns with insurance reimbursement (27/355, 7.6%). The older groups (aged over 65 years) showed greater concerns with regard to the operation of electronic devices (144/166, 86.7%; *P*<.001) as compared with the other groups (Figure 2). Those aged 35 to 49 years showed more concerns about the authenticity and reliability of e-hospitals (16/46, 34.8%; *P*=.046). Furthermore, young participants (aged 18-34 years) most often reported that they were reluctant to use e-hospitals because they were accustomed to face-to-face treatment (39/52, 75.0%; *P*<.001) and were concerned about insurance reimbursement (13/52, 25.0%; *P*<.001). Lastly, there was no age-related difference when considering “unhelpful for my specific disease” (*P*=.39).

Figure 1. Perceived facilitators for users (%).

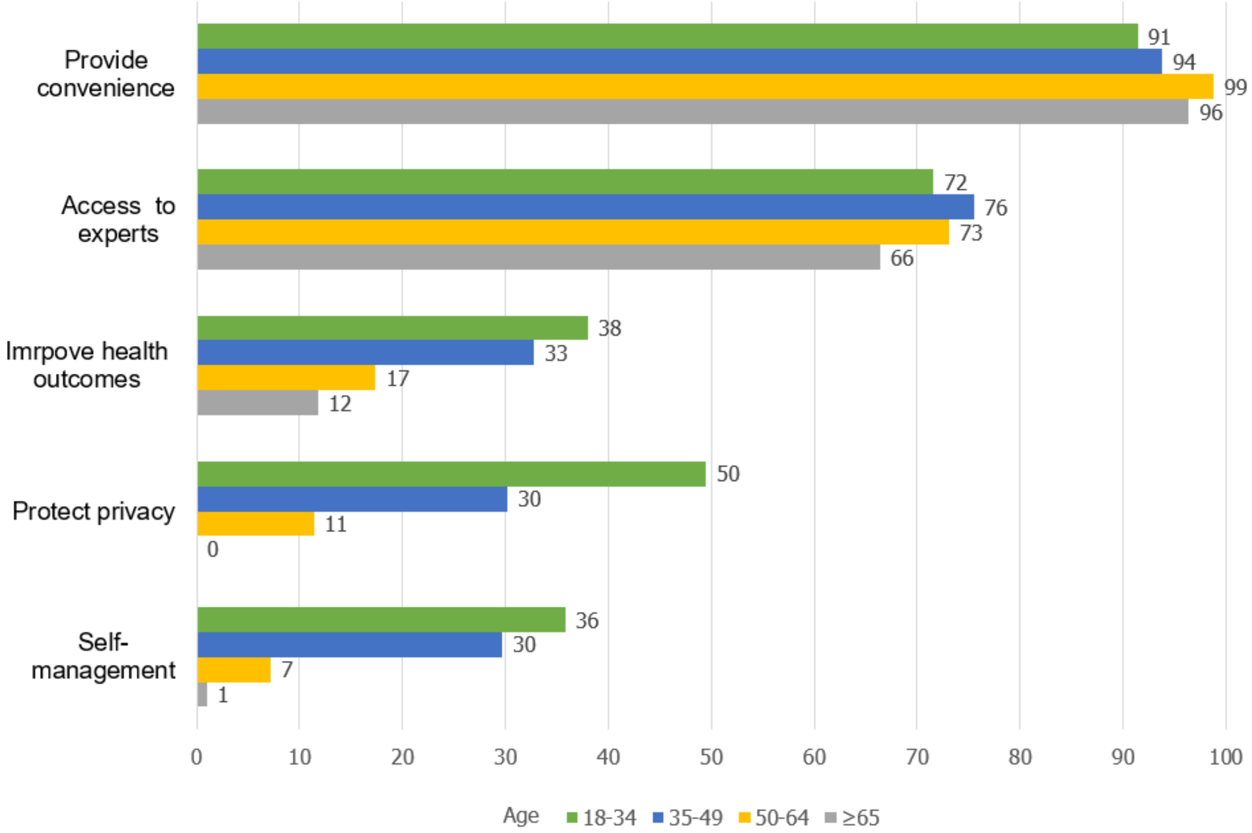
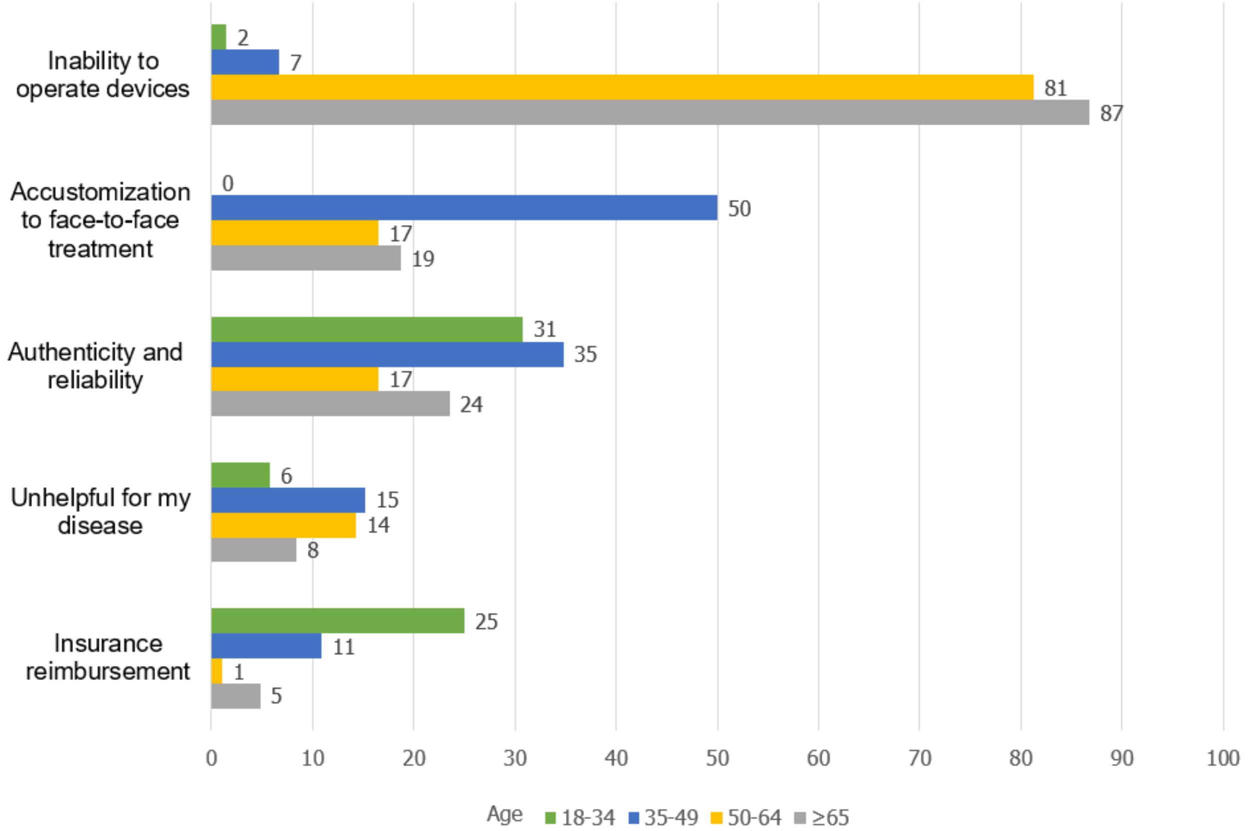


Figure 2. Perceived barriers for non-users (%).



Discussion

Study Importance

Research has shown that eHealth programs hold promise for improving health care accessibility and quality [50]; however, this success is contingent on patient acceptance. Therefore, the factors impacting the extent to which people are willing to engage with new eHealth interventions need to be understood in order to facilitate the adoption of these technologies [51,52]. This is the first survey to assess patients' attitudes, facilitators, and barriers toward e-hospital adoption and to provide related insights for decision makers about the essential factors necessary for successful adoption of e-hospitals in China's health care system.

Principal Findings: Part 1

Our results demonstrated that few patients in China know about e-hospitals and even fewer have experience using them. The proportion identified as familiar with e-hospitals (22.2%) in this study is consistent with what has been previously found for this region (18.6%) [53]. Both the prior study and our study indicate that urgent actions should be carried out to ensure broader awareness and familiarity with e-hospitals.

The health care needs and expectations of the older population have reached an unprecedented high in China as a result of reduced mobility and increasing morbidity [54]. In light of this, e-hospitals were theorized as a solution for increasing the efficiency of health care among elderly patients. Although previous research has shown that younger individuals are more likely to use web-based health tools [55,56], our results indicate that this age disparity regarding e-hospital readiness is based on technological proficiency. This may be explained by the pervasiveness of electronic devices in the modern age, which has caused young people to be more technologically proficient as compared with other generations. Nonetheless, this indicates that individuals ordinarily willing to use mobile health, such as elderly patients, may be deterred due to concerns regarding their ability to use technological devices [57,58]. Therefore, even though there is a growing number of e-hospital apps available to support elderly patients in China, their effectivity will likely be minimal unless such apps are designed to be user friendly and accessible for elderly populations.

In addition, our findings revealed that elderly people living with younger generations have higher usage of e-hospitals. This may be related to the fact that seniors living with children are able to get additional instructions and help from relatives. This finding is important, since in recent years, the one-child policy has contributed to reduced family size and an increase in the number of elderly "empty nesters" who are childless or whose children have already left home [59]. As a result, empty nesters' adoption will be a critical issue given that they are less likely to have access to technical support [60]. Therefore, it would be beneficial to provide solutions, such as offering e-hospital technical services at hospitals and over the internet, to facilitate e-hospital adoption for those who have minimal technological skills [61].

We found that educational disparity is an important determinant of the adoption of e-hospitals, which echoes the findings of previous research [62]. This finding once again underlines the importance of user-friendly interfaces and suggests that app developers should take the needs of people with any education level into account to further minimize the "digital gap" between users and nonusers [63].

Our results also indicated that working participants were more willing to use e-hospitals. This is unsurprising due to the time constraints work imposes on individuals and the convenience of e-hospitals. This is also in accordance with the results of previous studies that found employed individuals to be more interested in eHealth [62]. This indicates that modifying the opening hours of e-hospitals to better accommodate employed patients may play an important role in facilitating adoption.

Furthermore, we found that patients who had previous experiences with web-based medical services expressed positive interest toward e-hospital services. Our findings are in line with the results of previously published surveys, which found correlations between positive attitudes toward web-based health services and interest in eHealth in other contexts [64,65]. However, we found that patients requiring long-term disease management and postoperative rehabilitation did not have much higher willingness to use e-hospitals. This reluctance is congruent with the results of a previous survey that reported low interest toward telemedicine among patients with chronic lung disease [66]. One possibility is that patients with long-standing diseases or poor health conditions are less likely to trust a new less-mature health care delivery model, since they are more aware of the importance of regular thorough self-examination and follow-up. Regardless, these results were not expected, and they contradict one of the aims of e-hospitals, which is to heighten the efficiency and accessibility of chronic disease management and postoperative rehabilitative services. Given that there existed a positive correlation between previous experience with web-based health services and willingness to use e-hospitals, we suggest offering patients who have chronic illnesses or have undergone surgery a cost-free opportunity to get acquainted with e-hospital apps in order to encourage their usage and improve the success of this aim [67].

Principal Findings: Part 2

Research has shown that receptive patients are most encouraged by the convenience brought about by telemedicine [68], which is consistent with our findings and the current priority of e-hospitals.

However, there was less confidence regarding the "improve health outcomes" facilitator, given that a number of patients reported resistance to e-hospitals due to unassured authenticity and reliability of e-hospitals. Creating an effective e-hospital and integrating it into the system of physical hospitals is a complex endeavor that requires a multipronged strategy that addresses technical and operational issues constrained by local factors. For instance, maintaining the quality of nonprofit medical services is particularly challenging for e-hospitals constructed by for-profit companies. Therefore, establishing a government supervision mechanism [69] and increasing communication among stakeholders may play a key role in assuring the quality of adoption [70]. In addition, evaluation efforts should incorporate robust measures to document the outcomes of e-hospitals and establish a pathway for quick resolution of reported issues.

Given that privacy protection was one of the prominent facilitators for the use of e-hospitals among young individuals, cybersecurity must be prioritized in e-hospitals in order to prevent reluctance to adoption among patients [71,72]. This is substantiated by research [73] reporting that patients were worried about privacy protection when physicians used mobile devices and were particularly concerned with the possibility of individual health data and personal information being exposed. Therefore, we recommend that e-hospitals invest in cybersecurity capability development [74].

Moreover, insurance reimbursement is a key barrier, yet it has not been addressed by any of the 158 e-hospitals in China. In a study conducted in New York, it was reported that nearly half of users stopped using health care apps because of extra costs [46]. To avoid a similar situation, China's National Healthcare Security Administration should accelerate the pace of integrating e-hospitals into the public medical insurance reimbursement scheme [75].

Strengths, Limitations, and Future Research

This study has several strengths. First, the sample was relatively large and widely representative, which provided the opportunity for accurate examination of potential variations. Second, participants in the study were very sociodemographically and medically diverse, which allowed for in-depth analysis of variations in attitudes toward e-hospitals among a broad range of patients.

Despite the strengths, this study has several limitations. First, our sampling method was not randomized, and we used convenience sampling, such that participants were patients who happened to visit the hospitals of interest during the period of the survey. Consequently, our results are at risk of having statistical bias. Second, we did not collect text explanations in the survey, which thereby limited the possibility to investigate other facilitators and barriers to the adoption of e-hospitals. Third, as this was a pilot study, our results may not be generalizable to other regions with different sociodemographic characteristics, since we only included patients within one specific context of West China. However, it provided a preliminary analysis, and more research is needed to understand the complexity of the adoption of e-hospitals in China and elsewhere. Additional research will need to be conducted to better understand attitudes toward e-hospitals in other regions. Another limitation is the noninclusion of patients who underwent emergency procedures, and thus, our results do not take into account the possible intricacies of this cohort. Furthermore, this study does not provide insights into the attitudes of those who chose to not participate or complete the questionnaire; therefore, the specific needs and barriers of these users are unknown. We also did not record the demographics of nonresponders. Moreover, we measured connectivity to the internet using Wi-Fi, which is a common approach in the region; however, other forms of connectivity (eg, direct internet access through SIM cards) should be studied. It is hence plausible that there were limitations with the instrument and the collected data.

The integration of e-hospitals into China’s health care system requires an adequate understanding of not only patient attitudes toward acceptance but also the effects of such interventions. Although we identified several themes that can guide the research and technological development of e-hospitals, this study did not elucidate the potential risks of the application of this new technology. To maximize acceptability and usability, future research should focus on user testing with specific e-hospital prototypes. Such user testing should consider including a wider sample of the population, as well as measure the health care outcomes of e-hospital use. In addition, cost-effectiveness analysis is encouraged. Lastly, since the involvement of multiple stakeholders is essential to achieve effective adoption, implementation, and maintenance for systems in practice [76,77], further studies should consider the perspectives of various stakeholders, such as physicians and nurses.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant: 71874115) and the Science & Technology Department of Sichuan Province, China (grant: 2018KZ0046 and 2017FZ0104). The funder had no role in the design of the study; collection, analysis, and interpretation of data; or writing of the manuscript. We sincerely thank the three medical institutions and all participating individuals for helping conduct the survey. Additionally, we would like to thank all those who shared their perspectives throughout this study. We would like to acknowledge Dr Li Ma, Dr Yan Jiang, and Dr Zi Liu for their clinical assistance with this research. We would also like to sincerely acknowledge Mr Guanhua Qing for his assistance in improving this article.

Authors' Contributions

PL conceived and designed the study and questionnaire, carried out the survey and statistical analysis, and drafted the manuscript. YL was involved in the design of the questionnaire and was in charge of data collection and entry. XY performed the statistical analysis and interpretation of the data. JW helped with the design of the questionnaire. EM and MSJ consulted with the analysis and interpretation of the results, contributed to further development of the analysis and content, and revised the manuscript for important intellectual content. WL was the principal designer of the study and was responsible for all results of the study, as well as review and approval of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey questionnaire for patients’ perceptions of barriers and facilitators to the adoption of e-hospitals.

[PDF File (Adobe PDF File), 321 KB - jmir_v22i6e17221_app1.pdf]

References

1. Dai J, Wang X, Ayala FJ. Medical Informatics and the "Three Long, One Short" Problem of Large Urban Hospitals in China. JAMA 2016 Jul 19;316(3):269-270. [doi: 10.1001/jama.2016.5487] [Medline: 27434438]
2. Yip WC, Hsiao WC, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. Lancet 2012 Mar 03;379(9818):833-842. [doi: 10.1016/S0140-6736(11)61880-1] [Medline: 22386036]
3. Li X, Lu J, Hu S, Cheng K, De Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 09;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
4. General Office of the State Council. 2015 Sep 08. Guidance for promoting the construction of hirarchical healthcare system URL: http://www.gov.cn/zhengce/content/2015-09/11/content_10158.htm [accessed 2015-09-11]
5. State Council People’s Republic of China. 2018 Dec 24. Report on the financial distribution in healthcare URL: http://www.mof.gov.cn/zhengwuxinxi/caizhengxinwen/201812/t20181224_3102786.htm [accessed 2018-12-24]
6. Zeng L, Li Y, Zhang L, Liu G, Zhang Y, Zhen S, et al. Guideline use behaviours and needs of primary care practitioners in China: a cross-sectional survey. BMJ Open 2017 Sep 03;7(9):e015379 [FREE Full text] [doi: 10.1136/bmjopen-2016-015379] [Medline: 28871009]
7. Du S, Cao Y, Zhou T, Setiawan A, Thandar M, Koy V, et al. The knowledge, ability, and skills of primary health care providers in SEANERN countries: a multi-national cross-sectional study. BMC Health Serv Res 2019 Aug 27;19(1):602 [FREE Full text] [doi: 10.1186/s12913-019-4402-9] [Medline: 31455377]
8. Wang HH, Wang JJ, Zhou ZH, Wang XW, Xu L. General practice education and training in southern China: recent development and ongoing challenges under the health care reform. Malays Fam Physician 2013;8(3):2-10 [FREE Full text] [Medline: 25893051]
9. Yuan B, Balabanova D, Gao J, Tang S, Guo Y. Strengthening public health services to achieve universal health coverage in China. BMJ 2019 Jun 21;365:l2358 [FREE Full text] [doi: 10.1136/bmj.l2358] [Medline: 31227480]

10. Hesse BW, Greenberg AJ, Rutten LJ. The role of Internet resources in clinical oncology: promises and challenges. Nat Rev Clin Oncol 2016 Dec;13(12):767-776. [doi: 10.1038/nrclinonc.2016.78] [Medline: 27273045]
11. Gartner D, Padman R. E-HOSPITAL - A Digital Workbench for Hospital Operations and Services Planning Using Information Technology and Algebraic Languages. Stud Health Technol Inform 2017;245:84-88. [Medline: 29295057]
12. Hartzband P, Groopman J. Untangling the Web--patients, doctors, and the Internet. N Engl J Med 2010 Mar 25;362(12):1063-1066. [doi: 10.1056/NEJMp0911938] [Medline: 20335581]
13. Agarwal P, Kithulegoda N, Umpierre R, Pawlovich J, Pfeil JN, D'Avila OP, et al. Telemedicine in the driver's seat: new role for primary care access in Brazil and Canada: The Besrour Papers: a series on the state of family medicine in Canada and Brazil. Can Fam Physician 2020 Feb;66(2):104-111 [FREE Full text] [Medline: 32060190]
14. WHO Global Observatory for eHealth. Building foundations for eHealth: progress of Member States: report of the WHO Global Observatory for eHealth. Geneva, Switzerland: World Health Organization; Apr 09, 2007.
15. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the Next-Generation Perinatal Care: An Overview of the Literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
16. Jung H, Lee J. The impact of community-based eHealth self-management intervention among elderly living alone with hypertension. J Telemed Telecare 2017 Jan;23(1):167-173. [doi: 10.1177/1357633X15621467] [Medline: 26678063]
17. Das A, Faxvaag A, Svanæs D. The Impact of an eHealth Portal on Health Care Professionals' Interaction with Patients: Qualitative Study. J Med Internet Res 2015 Nov 24;17(11):e267 [FREE Full text] [doi: 10.2196/jmir.4950] [Medline: 26601678]
18. van der Meij E, Anema JR, Leclercq WK, Bongers MY, Consten EC, Schraffordt Koops SE, et al. Personalised perioperative care by e-health after intermediate-grade abdominal surgery: a multicentre, single-blind, randomised, placebo-controlled trial. Lancet 2018 Jul 07;392(10141):51-59. [doi: 10.1016/S0140-6736(18)31113-9] [Medline: 29937195]
19. Lu C, Hu Y, Xie J, Fu Q, Leigh I, Governor S, et al. The Use of Mobile Health Applications to Improve Patient Experience: Cross-Sectional Study in Chinese Public Hospitals. JMIR Mhealth Uhealth 2018 May 23;6(5):e126 [FREE Full text] [doi: 10.2196/mhealth.9145] [Medline: 29792290]
20. State Council People’s Republic of China. 2015 Jul 01. Opinions on actively promoting internet projects URL: http://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm [accessed 2015-07-04]
21. Huckvale K, Wang CJ, Majeed A, Car J. Digital health at fifteen: more human (more needed). BMC Med 2019 Mar 18;17(1):62 [FREE Full text] [doi: 10.1186/s12916-019-1302-0] [Medline: 30879466]
22. National Health and Family Planning Commission. 2018 Aug 14. Opinions on the development of e-hospital URL: http://www.gov.cn/xinwen/2018-09/14/content_5322040.htm [accessed 2018-08-14]
23. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015 Aug;3(8):e445-e446 [FREE Full text] [doi: 10.1016/S2214-109X(15)00042-X] [Medline: 26187488]
24. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet Hospitals in China: Cross-Sectional Survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
25. Qiu Y, Liu Y, Ren W, Qiu Y, Ren J. Internet-Based and Mobile-Based General Practice: Cross-Sectional Survey. J Med Internet Res 2018 Sep 25;20(9):e266 [FREE Full text] [doi: 10.2196/jmir.8378] [Medline: 30257819]
26. National Bureau of Statistics of China. China Health Statistics Yearbook 2018. Beijing: China Statistics Publishing House; 2018.
27. Tencent. 2019 Mar 10. Tengai doctor service closure notice URL: https://doctor.qq.com [accessed 2019-03-10]
28. Tencent. 2017 Apr 01. Baidu doctor service closure notice URL: https://finance.qq.com/a/20170303/027165.htm [accessed 2017-04-01]
29. Keen P. Shaping the Future. Business Design through Information Technology. In: Journal of Information Technology. Boston: Havard Business School; Sep 1991:237-237.
30. Kijsanayotin B, Pannarunothai S, Speedie SM. Factors influencing health information technology adoption in Thailand's community health centers: applying the UTAUT model. Int J Med Inform 2009 Jun;78(6):404-416. [doi: 10.1016/j.ijmedinf.2008.12.005] [Medline: 19196548]
31. Saleh S, Khodor R, Alameddine M, Baroud M. Readiness of healthcare providers for eHealth: the case from primary healthcare centers in Lebanon. BMC Health Serv Res 2016 Nov 10;16(1):644 [FREE Full text] [doi: 10.1186/s12913-016-1896-2] [Medline: 27832788]
32. Li P, Luo Y, Yu X, Zeng Z, Jin W, Mason E, et al. Readiness of Healthcare Providers for e-Hospitals: Cross-sectional Analysis in China. SSRN Journal 2020 Apr 20:1-10. [doi: 10.2139/ssrn.3573282]
33. Khoja S, Scott RE, Casebeer AL, Mohsin M, Ishaq A, Gilani S. e-Health readiness assessment tools for healthcare institutions in developing countries. Telemed J E Health 2007 Aug;13(4):425-431. [doi: 10.1089/tmj.2006.0064] [Medline: 17848110]
34. Khoja S, Scott R, Gilani S. E-health readiness assessment: promoting "hope" in the health-care institutions of Pakistan. World Hosp Health Serv 2008;44(1):36-38. [Medline: 18549033]
35. Kiberu VM, Mars M, Scott RE. Barriers and opportunities to implementation of sustainable e-Health programmes in Uganda: A literature review. Afr J Prim Health Care Fam Med 2017 May 29;9(1):e1-e10 [FREE Full text] [doi: 10.4102/phcfm.v9i1.1277] [Medline: 28582996]

36. Yu P, Li H, Gagnon M. Health IT acceptance factors in long-term care facilities: a cross-sectional survey. Int J Med Inform 2009 Apr;78(4):219-229. [doi: 10.1016/j.ijmedinf.2008.07.006] [Medline: 18768345]
37. Durrani H, Khoja S, Naseem A, Scott R, Gul A, Jan R. Health needs and eHealth readiness assessment of health care organizations in Kabul and Bamyan, Afghanistan. East Mediterr Health J 2012 Jun;18(6):663-670 [FREE Full text] [doi: 10.26719/2012.18.6.663] [Medline: 22888626]
38. Ruiz Morilla MD, Sans M, Casasa A, Giménez N. Implementing technology in healthcare: insights from physicians. BMC Med Inform Decis Mak 2017 Jun 27;17(1):92 [FREE Full text] [doi: 10.1186/s12911-017-0489-2] [Medline: 28655299]
39. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. BMC Health Serv Res 2018 Mar 27;18(1):220 [FREE Full text] [doi: 10.1186/s12913-018-2986-0] [Medline: 29587721]
40. Nahm E, Blum K, Scharf B, Friedmann E, Thomas S, Jones D, et al. Exploration of Patients' Readiness for an eHealth Management Program for Chronic Heart Failure. The Journal of Cardiovascular Nursing 2008;23(6):463-471. [doi: 10.1097/01.jcn.0000317459.41015.d6]
41. Muigg D, Kastner P, Modre-Osprian R, Haluza D, Duftschmid G. Is Austria Ready for Telemonitoring? A Readiness Assessment Among Doctors and Patients in the Field of Diabetes. Stud Health Technol Inform 2018;248:322-329. [Medline: 29726454]
42. Schwarz F, Ward J, Willcock S. E-Health readiness in outback communities: an exploratory study. Rural Remote Health 2014;14(3):2871 [FREE Full text] [Medline: 25190566]
43. Katz SJ, Moyer CA. The emerging role of online communication between patients and their providers. J Gen Intern Med 2004 Sep;19(9):978-983 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30432.x] [Medline: 15333064]
44. Statistical Bureau Of Sichuan. NBS survey office in Sichuan: Sichuan Statistical Yearbook. Beijing: China Statistics Press; 2018.
45. Tao W, Zeng W, Yan L, Yang H, Wen J, Li W. The health service capacity of primary health care in West China: different perspectives of physicians and their patients. BMC Health Serv Res 2019 Feb 28;19(1):143 [FREE Full text] [doi: 10.1186/s12913-019-3964-x] [Medline: 30819168]
46. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
47. Albrecht U, Afshar K, Illiger K, Becker S, Hartz T, Breil B, et al. Expectancy, usage and acceptance by general practitioners and patients: exploratory results from a study in the German outpatient sector. Digit Health 2017;3:2055207617695135 [FREE Full text] [doi: 10.1177/2055207617695135] [Medline: 29942582]
48. Meng Q, Mills A, Wang L, Han Q. What can we learn from China's health system reform? BMJ 2019 Jun 19;365:l2349 [FREE Full text] [doi: 10.1136/bmj.l2349] [Medline: 31217222]
49. Chengdu Bureau of Statistics. Announcement on the average salary of all employed persons in 2017. Beijing: China Statistics Press; Jan 01, 2018.
50. Melchiorre MG, Lamura G, Barbabella F, ICARE4EU Consortium. eHealth for people with multimorbidity: Results from the ICARE4EU project and insights from the "10 e's" by Gunther Eysenbach. PLoS One 2018;13(11):e0207292 [FREE Full text] [doi: 10.1371/journal.pone.0207292] [Medline: 30427924]
51. AlBar AM, Hoque MR. Patient Acceptance of e-Health Services in Saudi Arabia: An Integrative Perspective. Telemed J E Health 2019 Sep;25(9):847-852. [doi: 10.1089/tmj.2018.0107] [Medline: 30452333]
52. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and Facilitators of Engagement With mHealth Technology for Remote Measurement and Management of Depression: Qualitative Analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
53. Chen P, Xiao L, Gou Z, Xiang L, Zhang X, Feng P. Telehealth attitudes and use among medical professionals, medical students and patients in China: A cross-sectional survey. Int J Med Inform 2017 Dec;108:13-21. [doi: 10.1016/j.ijmedinf.2017.09.009] [Medline: 29132618]
54. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. The Lancet 2019 Sep;394(10204):1192-1204. [doi: 10.1016/s0140-6736(19)32136-1]
55. de Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
56. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2016 Oct 16;24(1):4-12. [doi: 10.1177/1357633x16674087]
57. Lee J, Nguyen AL, Berg J, Amin A, Bachman M, Guo Y, et al. Attitudes and preferences on the use of mobile health technology and health games for self-management: interviews with older adults on anticoagulation therapy. JMIR Mhealth Uhealth 2014 Jul 23;2(3):e32 [FREE Full text] [doi: 10.2196/mhealth.3196] [Medline: 25098413]
58. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and Acceptance of mHealth in Patients With Cardiovascular Diseases: A Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]

59. Qian Y, Qin W, Zhou C, Ge D, Zhang L, Sun L. Utilisation willingness for institutional care by the elderly: a comparative study of empty nesters and non-empty nesters in Shandong, China. BMJ Open 2018 Aug 17;8(8):e022324 [FREE Full text] [doi: 10.1136/bmjopen-2018-022324] [Medline: 30121610]
60. Zhang C, Xue Y, Zhao H, Zheng X, Zhu R, Du Y, et al. Prevalence and related influencing factors of depressive symptoms among empty-nest elderly in Shanxi, China. J Affect Disord 2019 Feb 15;245:750-756. [doi: 10.1016/j.jad.2018.11.045] [Medline: 30448759]
61. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, Implementing, and Evaluating Mobile Health Technologies for Managing Chronic Conditions in Older Adults: A Scoping Review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
62. Sun L, Wang Y, Greene B, Xiao Q, Jiao C, Ji M, et al. Facilitators and barriers to using physical activity smartphone apps among Chinese patients with chronic diseases. BMC Med Inform Decis Mak 2017 Apr 19;17(1):44 [FREE Full text] [doi: 10.1186/s12911-017-0446-0] [Medline: 28420355]
63. Bailey SC, O'Conor R, Bojarski EA, Mullen R, Patzer RE, Vicencio D, et al. Literacy disparities in patient access and health-related use of Internet and mobile technologies. Health Expect 2015 Dec;18(6):3079-3087 [FREE Full text] [doi: 10.1111/hex.12294] [Medline: 25363660]
64. Hofstede J, de Bie J, van Wijngaarden B, Heijmans M. Knowledge, use and attitude toward eHealth among patients with chronic lung diseases. Int J Med Inform 2014 Dec;83(12):967-974. [doi: 10.1016/j.ijmedinf.2014.08.011] [Medline: 25269992]
65. Rising CJ, Bol N, Kreps GL. Age-Related Use and Perceptions of eHealth in Men With Prostate Cancer: A Web-Based Survey. JMIR Cancer 2015 Jun 25;1(1):e6 [FREE Full text] [doi: 10.2196/cancer.4178] [Medline: 28410165]
66. Duplaga M. The acceptance of e-health solutions among patients with chronic respiratory conditions. Telemed J E Health 2013 Sep;19(9):683-691 [FREE Full text] [doi: 10.1089/tmj.2012.0306] [Medline: 23734700]
67. Cranen K, Veld RH, Ijzerman M, Vollenbroek-Hutten M. Change of patients' perceptions of telemedicine after brief use. Telemed J E Health 2011 Sep;17(7):530-535. [doi: 10.1089/tmj.2010.0208] [Medline: 21767150]
68. Valikodath NG, Leveque TK, Wang SY, Lee PP, Newman-Casey PA, Hansen SO, et al. Patient Attitudes Toward Telemedicine for Diabetic Retinopathy. Telemed J E Health 2017 Mar;23(3):205-212 [FREE Full text] [doi: 10.1089/tmj.2016.0108] [Medline: 27336678]
69. Snowdon DA, Hau R, Leggat SG, Taylor NF. Does clinical supervision of health professionals improve patient safety? A systematic review and meta-analysis. Int J Qual Health Care 2016 Sep;28(4):447-455. [doi: 10.1093/intqhc/mzw059] [Medline: 27283436]
70. Jalali M, Rahmandad H, Bullock SL, Lee-Kwan SH, Gittelsohn J, Ammerman A. Dynamics of intervention adoption, implementation, and maintenance inside organizations: The case of an obesity prevention initiative. Soc Sci Med 2019 Mar;224:67-76. [doi: 10.1016/j.socscimed.2018.12.021] [Medline: 30763824]
71. Jalali MS, Razak S, Gordon W, Perakslis E, Madnick S. Health Care and Cybersecurity: Bibliometric Analysis of the Literature. J Med Internet Res 2019 Feb 15;21(2):e12644 [FREE Full text] [doi: 10.2196/12644] [Medline: 30767908]
72. Jalali MS, Kaiser JP. Cybersecurity in Hospitals: A Systematic, Organizational Perspective. J Med Internet Res 2018 May 28;20(5):e10059 [FREE Full text] [doi: 10.2196/10059] [Medline: 29807882]
73. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht U. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
74. Jalali MS, Siegel M, Madnick S. Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. The Journal of Strategic Information Systems 2019 Mar;28(1):66-82. [doi: 10.1016/j.jsis.2018.09.003]
75. National Healthcare Security Administration. 2019 Aug 30. Opinions on the price of internet medical services and medical insurance payment policies URL: http://www.nhsa.gov.cn/art/2019/8/30/art_37_1707.html [accessed 2019-08-30]
76. Craven MP, Lang AR, Martin JL. Developing mHealth Apps with Researchers: Multi-Stakeholder Design Considerations. In: Design, User Experience, and Usability. User Experience Design for Everyday Life Applications and Services. 2014 Presented at: DUXU 2014; June 22-27, 2014; Heraklion, Crete, Greece. [doi: 10.1007/978-3-319-07635-5_2]
77. Jalali M, Rahmandad H, Bullock S, Ammerman A. Dynamics of Implementation and Maintenance of Organizational Health Interventions. Int J Environ Res Public Health 2017 Aug 15;14(8) [FREE Full text] [doi: 10.3390/ijerph14080917] [Medline: 28809807]

Abbreviations

eHealth: electronic health
ICT: information and communication technology
IT: information technology
PHC: primary health care center

Edited by G Eysenbach; submitted 26.11.19; peer-reviewed by M Balikuddembe, R Scott, D Gunasekeran, Z Ma; comments to author 20.01.20; revised version received 11.03.20; accepted 12.04.20; published 11.06.20.

Please cite as:
Li P, Luo Y, Yu X, Wen J, Mason E, Li W, Jalali MS
Patients' Perceptions of Barriers and Facilitators to the Adoption of E-Hospitals: Cross-Sectional Study in Western China
J Med Internet Res 2020;22(6):e17221
URL: http://www.jmir.org/2020/6/e17221/
doi:10.2196/17221
PMID:32525483

Original Paper

An Intervention Delivered by Mobile Phone Instant Messaging to Increase Acceptability and Use of Effective Contraception Among Young Women in Bolivia: Randomized Controlled Trial

Ona L McCarthy[1], BA, MSc, PhD; Carolina Aliaga[2], BA; Maria Eugenia Torrico Palacios[2], MSc; Jhonny López Gallardo[2], MD; Silvia Huaynoca[3], MPH, MD; Baptiste Leurent[1], MSc, PhD; Phil Edwards[1], PhD; Melissa Palmer[1], BSc, MSc, PhD; Irrfan Ahamed[1], BSc; Caroline Free[1], PhD, MBChB

[1]The London School of Hygiene and Tropical Medicine, London, United Kingdom

[2]CIES Salud Sexual Salud Reproductiva, La Paz, Bolivia

[3]International Planned Parenthood Federation/Western Hemisphere Region, New York, NY, United States

Corresponding Author:
Ona L McCarthy, BA, MSc, PhD
The London School of Hygiene and Tropical Medicine
Keppel Street
London
United Kingdom
Phone: 44 79272581
Email: ona.mccarthy@lshtm.ac.uk

Abstract

Background: Although the most effective methods of contraception are available in Bolivia, unmet need for contraception among women aged 15 to 19 years is estimated to be 38% (2008), and the adolescent fertility rate is 71 per 1000 women (2016). Mobile phones are a popular mode to deliver health behavior support. We developed a contraceptive behavioral intervention for young Bolivian women delivered by mobile phone and guided by behavioral science. The intervention consists of short instant messages sent through an app over 4 months.

Objective: This trial aimed to evaluate the effect of the intervention on young Bolivian women’s use of and attitudes toward the effective contraceptive methods available in Bolivia.

Methods: This was a parallel group, individually randomized superiority trial with a 1:1 allocation ratio. Women were eligible if they were aged 16 to 24 years, owned a personal Android mobile phone, lived in La Paz or El Alto, reported an unmet need for contraception, and could read Spanish. The target sample size was 1310 participants. Participants allocated to the intervention had access to an app with standard family planning information and intervention messages. Participants allocated to the control group had access to the same app and control messages. Coprimary outcomes were use of effective contraception and acceptability of at least one method of effective contraception at 4 months. Secondary outcomes were use of effective contraception during the study, acceptability of the individual methods, service uptake, unintended pregnancy, and abortion. Process outcomes included knowledge, perceived norms, personal agency, and intention. Outcomes were analyzed using logistic and linear regression. We also asked participants about physical violence.

Results: A total of 640 participants were enrolled, and 67.0% (429) of them contributed follow-up data for the coprimary outcome, the use of effective contraception. There was no evidence that use differed between the groups (33% control vs 37% intervention; adjusted odds ratio [OR] 1.19, 95% CI 0.80 to 1.77; *P*=.40). There was a borderline significant effect regarding acceptability (63% control vs 72% intervention; adjusted OR 1.49, 95% CI 0.98 to 2.28; *P*=.06). There were no statistically significant differences in any of the secondary or process outcomes. The intervention dose received was low. In the control group, 2.8% (6/207) reported experiencing physical violence compared with 1.9% (4/202) in the intervention group (Fisher exact test *P*=.75).

Conclusions: This trial was unable to provide definitive conclusions regarding the effect of the intervention on use and acceptability of effective contraception because of under recruitment. Although we cannot strongly recommend implementation, the results suggest that it would be safe and may increase the acceptability of effective contraception if the intervention messages were offered alongside the download of the app.

Trial Registration: ClinicalTrials.gov NCT02905526; https://clinicaltrials.gov/ct2/show/NCT02905526

(J Med Internet Res 2020;22(6):e14073) doi:10.2196/14073

KEYWORDS

Bolivia; contraception; mobile phone; cellphone; reproductive health; young adults

Introduction

Unintended pregnancy is associated with numerous poorer health outcomes for both women and their children [1-5]. Satisfying the unmet need for contraception is essential in helping women avoid unintended pregnancies, which requires an understanding of the barriers to use in specific settings [6]. A woman who has an unmet need for contraception is of reproductive age (15-49 years); is legally married, cohabiting, in a consensual union, or unmarried and sexually active; is not using any method of contraception; and is fecund and does not want to have a child (or another child) in the next 2 years or at all [6]. Modern contraceptive methods include oral contraceptives, injectables, intrauterine devices (IUDs), implants, the patch, the ring, male and female sterilization, male and female condoms and other barrier methods, modern fertility awareness methods, and emergency contraception [7]. *Effective* contraceptive methods are methods with less than 10% typical use failure rate at 12 months, that is, all modern methods besides condoms, other barrier methods, and modern fertility awareness methods [8,9].

In Bolivia, the latest Demographic and Health Survey (2008) reported that 85% of sexually active women aged between 15 and 19 years wanted to avoid a pregnancy, yet only 49% of these women reported using any contraceptive method [10]. Male condoms and the injection were the most common modern methods reported by this group (19.6% and 6.2%, respectively), with 2% reporting that they use withdrawal and 13% using periodic abstinence. A more recent survey (2016) reported that the adolescent fertility rate was 71 per 1000 women [11]. The 2016 survey found that among unmarried, sexually active women aged 15 to 19 years, an estimated 34% were not using any method of contraception [11], which was down from 52% in the 2008 survey [10]. Although unmet need may have changed since, the most recent data from 2008 suggest that unmet need among women aged 15 to 19 years was estimated to be 38% [10,12]. The (nonpermanent) effective methods available in Bolivia are oral contraceptive pills, IUDs, injectables, implants, and the patch. These methods are available for a fee (between US $1.50 and US $3.00) at the Centro de Investigación, Educación y Servicios (CIES), our partner organization, which operates 18 clinics in 7 of the 9 departments in the country. The public health care system provides only condoms to young people; however, young people usually buy condoms from private pharmacies.

The option of delivering health interventions by a mobile phone has gained popularity, particularly over the last decade [13-24]. Using mobile phones to deliver support to young people regarding sexual and reproductive health at the time of their choosing may be particularly useful, given the sensitivity of the topic. Randomized controlled trials have provided evidence that interventions delivered by a mobile phone can improve contraceptive use [25-27] and knowledge [25,28-30]. Other trials, however, have not found a beneficial effect [31-35]. The mixed evidence could be because of variability in the quality of the intervention development (eg, the target group may not have been adequately consulted), the content of the intervention (eg, the intervention content was not grounded in theory or behavioral science), and trial methodology.

The London School of Hygiene & Tropical Medicine (LSHTM) and CIES developed a contraceptive behavioral intervention for young Bolivian women delivered by mobile phone [36]. We developed the intervention guided by an established approach grounded in behavioral science [37]. The intervention is informed by the Integrated Behavioral Model (IBM) [38] and consists of short instant messages sent through CIES's *Tú decides* app over 4 months. In this report, we present the results of the evaluation of the intervention by randomized controlled trial. The aim of the trial was to establish if the intervention increases young Bolivian women's use and acceptability of the effective contraceptive methods.

Methods

Study Design and Participants

The methods reported in this section were first published in the trial protocol [39] and the statistical analysis plan [40].This was a parallel group, individually randomized superiority trial with a 1:1 allocation ratio that evaluated the effect of the intervention delivered by CIES's app. Women were eligible to take part if they were aged 16 to 24 years, owned a personal Android mobile phone, lived in La Paz or El Alto, reported an unmet need for contraception (ie, are sexually active, not using effective contraception, and want to avoid a pregnancy), and could read Spanish. Participants must have been willing to download the app and receive messages about contraception on their mobile phone. The trial was promoted through CIES's service delivery points in La Paz and El Alto, the CIES website, flyers distributed through CIES's youth network, and social media sites. Participants provided informed consent through the secure web-based trial database and randomization system.

Ethics Approval and Consent to Participate

The trial was granted ethical approval by LSHTM Interventions Research Ethics Committee on May 16, 2016, and by La Comisión de Ética de la Investigación del Comité Nacional de Bioética on September 20, 2016.

Intervention and Control

The intervention was developed with young Bolivian people in 2015 to 2016, guided by a systematic protocol for the development of behavior change interventions [36]. The development process involved the following steps: (1) needs assessment (activities included establishing a project planning group, a literature search, focus group discussions, and interviews with the target group and interviews with local service providers), (2) specifying behavioral change to result from the intervention, (3) designing the intervention components by selecting behavior change methods, and (4) producing and refining the intervention content. The needs assessment revealed that young people in El Alto and La Paz were eager to receive information about contraception on their phone, lacked comprehensive knowledge about contraception, and expressed a range of negative beliefs about effective methods. The intervention messages were tested with young people and refined after each consultation in an iterative process until the context was acceptable to them.

The intervention provided accurate information about contraception, targeted the beliefs identified in the development phase that influence contraceptive use (eg, specific misconceptions about the side effects and health risks of contraception), and aimed to support young women in believing that they can influence their reproductive health. The messages contained 10 behavior change methods [41], adapted for delivery by mobile phone: belief selection, facilitation, anticipated regret, guided practice, verbal persuasion, tailoring, cultural similarity, arguments, shifting perspective, and goal setting. Participants allocated to the intervention group received 0 to 3 messages per day (a total of 183 messages) for 120 days. Please see the protocol [39] and the intervention development publication [36] for a more detailed description of the intervention.

The Tú decides app itself contained standard family planning information and no behavior change methods. Participants allocated to the intervention arm had access to the app and the intervention instant messages. Participants allocated to the control arm had access to the Tú decides app and 7 control instant messages about the importance of their participation and reminding them to contact the project coordinator if they change their number (which intervention participants also received). All participants who received usual care were free to seek any other support, whether existing or new.

Allocation and Intervention Delivery

After providing informed consent, participants completed the baseline questionnaire through the database and randomization system. The allocation sequence was generated by the remote computer-based randomization software. Randomization occurred immediately after baseline data were submitted. All participants downloaded the app immediately after they submitted their baseline data. The messages commenced within 24 hours after participants downloaded the app.

Protecting Against Bias

Owing to the nature of the intervention, participants would have been aware of the allocation soon after they started receiving the messages. Local research staff collecting outcome data were masked to allocation unless the participant revealed it to them. Researchers who analyzed the data were masked to treatment allocation.

Outcomes

Coprimary Outcomes

The coprimary outcomes at 4 months post randomization were (1) self-reported current use of effective contraception and (2) the proportion of participants reporting that at least one method of effective contraception was acceptable. The primary outcome measure was constructed based on guidelines for measuring IBM constructs [38,42,43] and tested for face validity with the target group. The acceptability of each method was measured by the following stems: “Using the [method] ...causes infertility, ...causes unwanted side effects, ...is easy, ...is a good way to prevent pregnancy” and “I would recommend the [method] to a friend.” The IUD and implant include an additional stem: “The [method] insertion would not be a problem for me.” The response options for each stem were “strongly disagree,” “disagree,” “not sure,” “agree,” “strongly agree,” and “I do not know what the [method] is.” A method was acceptable if participants reported “agree” or “strongly agree” for all scales except for “...causes infertility” and “...causes unwanted side effects” stems, for which “disagree” or “strongly disagree” indicated acceptability [39].

Secondary Outcomes

The following secondary outcomes were self-reported: use of effective contraception at any time during the study, acceptability of each effective contraception method, attendance at a sexual health service during the study, unintended pregnancy during the study (the proportion reporting that they became pregnant and they did not want to become pregnant), and abortion during the study.

Process Outcomes

The following process outcomes were self-reported: knowledge of effective contraception, perceived norms, and personal agency in relation to using and communicating with partners about contraception, intention to use effective contraception, and intervention dose received.

Data Collection

Data were collected at baseline and at 4 months post randomization using questionnaires. At baseline, we collected personal and demographic data and the coprimary outcome acceptability. At follow-up, we collected all outcomes and the following: if participants report using an effective method, where they obtained it; current pregnancy intention; whether they knew someone else who had also participated in the study and, if so, if they read each other’s messages; and to assess potential adverse outcomes, we asked participants if they have experienced physical violence since being in the study and if anything good or bad happened as a result of receiving the messages. An instant message that included a link to the database to complete the follow-up questionnaire was sent to all participants through the app 4 months after downloading the app. If participants did not complete the follow-up questionnaire themselves, local research staff unaware of participants’ allocation contacted them by telephone to collect their data.

Sample Size

The trial was powered to detect a 10% absolute difference in the use of effective contraception between the intervention and control groups at 4 months. The 10% difference was based on a trial evaluating a postabortion mobile phone intervention, which found that 18% more women in the intervention arm than those in the control arm were using effective contraception at 4 months (64% vs 46%) [27]. We assumed that our trial would observe a smaller increase in contraceptive uptake, as it does not specifically involve women who had just had an abortion, who had already accessed services, and who may also have a greater intention to use contraception compared with the women in our trial. Therefore, we powered our trial to detect a smaller absolute difference of 10% uptake in effective contraception at 4 months.

The best estimate at the time of designing the trial was that the proportion of women aged 16 to 24 years in a partnership living in La Paz or El Alto using effective contraception was around 44% [44]. In total, 1048 participants provided 90% power to detect a 10% absolute difference, at the 5% significance level, assuming 44% use in the control group versus 54% in the intervention group (corresponding to an odds ratio [OR] of 1.49). Allowing for 20% loss to follow-up, the sample size was 1310.

Statistical Analysis

The trial protocol was accepted for publication on November 3, 2017 [39], and the detailed statistical analysis plan was publicly released on November 7, 2017 [40]. Analyses were according to the randomized arm, and only participants with complete outcome data were included in the principal analysis. All statistical tests were two sided and were considered significant at the 5% level. The analysis was conducted using Stata 15 (StataCorp). Unmasking occurred on February 6, 2018, after the analyses outlined within the analysis plan were complete on masked data.

Loss to Follow-Up and Missing Data

We used a chi-square test to investigate the evidence for whether losses to follow-up differed by trial arm. We used logistic regression to compare the baseline characteristics of participants who completed follow-up with those who did not. We investigated whether predictors of loss to follow-up differed by trial arm by testing for an interaction.

Principal Analysis

Analysis of the Coprimary Outcomes

We compared the proportion that reported using a method of effective contraception or finding at least one method acceptable between the groups using logistic regression. We report the crude and adjusted OR along with the 95% CI and P value. The primary analysis was adjusted for the following prespecified baseline covariates: age (16-19 and 20-24 years), number of children (0 and ≥1), education level (university and other), and acceptability of effective contraception at baseline (at least one method acceptable and no methods acceptable) [39,40].

Analysis of the Secondary Outcomes

The analysis of the secondary outcomes was similar to the analysis of the primary outcome, although for the acceptability of the individual methods, the acceptability of that method at baseline replaced the acceptability of at least one method at baseline as a covariate.

Analysis of the Process Outcomes

The process outcomes comprised ordinal scales. Each scale was analyzed individually using ordered logistic regression to estimate proportional ORs. For knowledge, each correct answer received one point. The points were summed, and an overall score was produced for analysis. We used linear regression to test for a difference in the mean scores between the trial arms. To assess the *dose* of the intervention that the intervention participants received, we analyzed the number of messages that participants reported to have read (all, most, some, and none) and whether they stopped the messages.

Additional Analyses

Sensitivity Analyses

We conducted 2 sensitivity analyses allowing for the missing data. In the first analysis (an *extreme case* analysis), we considered that all participants lost to follow-up did not use an effective method of contraception or did not find at least one method acceptable. In the second analysis, we adjusted for the main baseline predictors of missingness. Both sensitivity analyses were adjusted for the baseline covariates, as mentioned above.

Subgroup Analysis

We conducted an exploratory subgroup analysis for each coprimary outcome to determine if the intervention effect varied by baseline characteristics. The prespecified subgroups were age (split at the median), marital status (married/not married), number of children (0/≥1), geographical location (El Alto/La Paz), occupation (in education/other), and education level (university/other). Within the subgroups, we assessed the heterogeneity of treatment effect with a test for interaction [45-49]. We estimated ORs with 95% CIs for each subgroup.

Contamination

To assess the potential for contamination, we report the proportion of control group participants who reported that they read another participant’s messages and the proportion of intervention participants who reported that their messages were read by another participant.

Report of Physical Violence

We report the proportion of participants in each group who reported experiencing physical violence during the study.

Results

Recruitment, Randomization, and Exclusions

Between March 1, 2017, and July 29, 2017, there were 645 randomizations by the system. Follow-up ended on February 8, 2018. During the trial follow-up, we discovered that 3 participants enrolled and were randomized twice. For the 1 participant who was allocated to the same arm on both randomizations, we kept this participant in the analysis using the baseline data from their first record. For the 2 participants who were allocated to different arms, we excluded them from the analysis. This resulted in 640 participants included in the trial.

Of the 640 participants, 321 were allocated to the intervention arm and 319 participants were allocated to the control arm (Figure 1).

Baseline Characteristics

The baseline characteristics of trial participants are reported in Table 1. Mean age was 20 years. In addition, 90.4% (579/640) of the participants did not have children, and only 8% (26/640) of the participants found at least one method of effective contraception acceptable. Acceptability was very low for the individual methods (1%-3%). Characteristics were similar between the 2 groups.

Figure 1. Trial flow diagram.

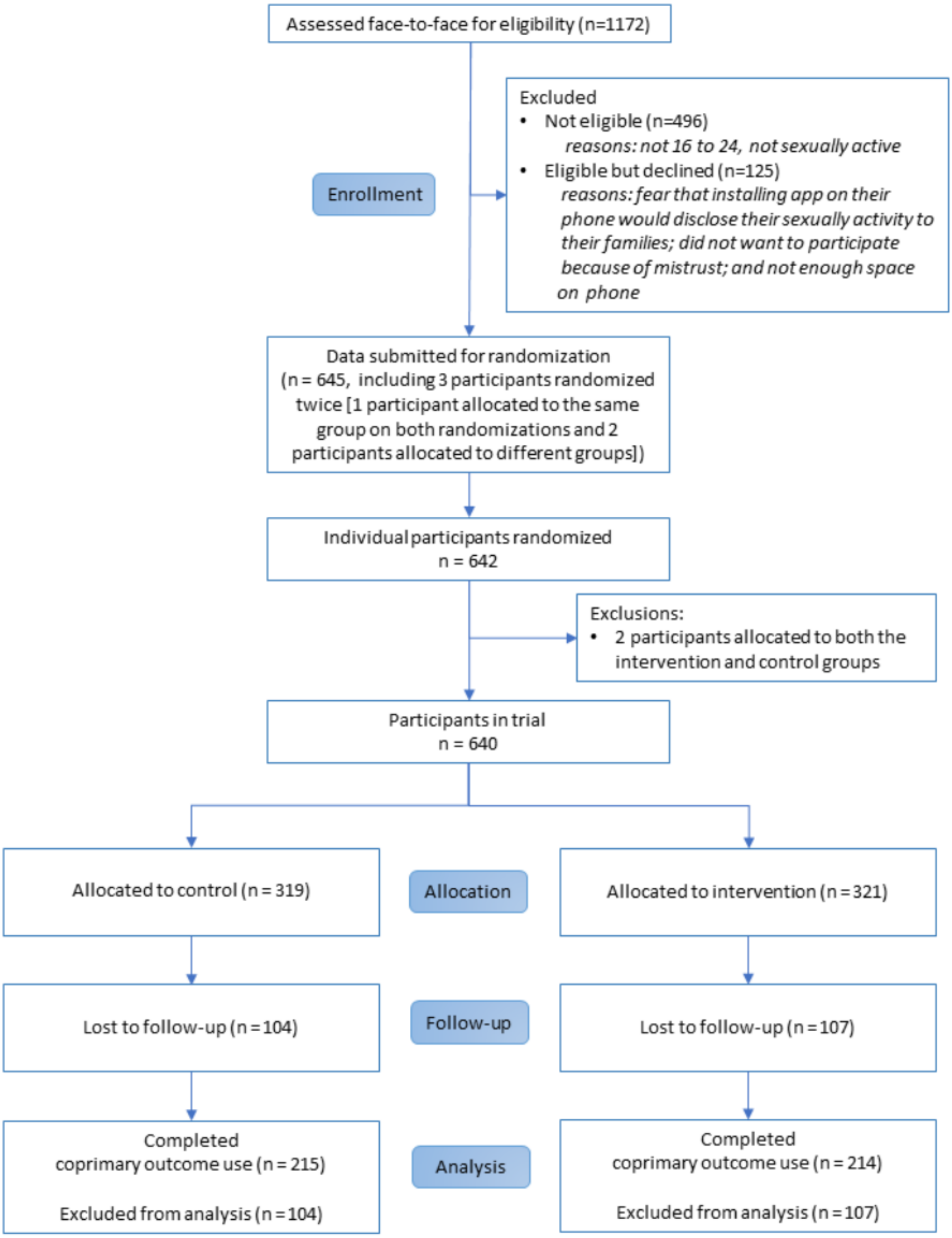


Table 1. Baseline characteristics.

Characteristic	Control (n=319)	Intervention (n=321)	All participants (N=640)
Age (years)			
Mean (SD)	20.42 (2.6)	20.27 (2.9)	20.35 (2.6)
Range, n (%)			
16-19	150 (47.0)	165 (51.4)	315 (49.2)
20-24	169 (52.9)	156 (48.6)	325 (50.8)
Marital status, n (%)			
Married	14 (4.4)	18 (5.6)	32 (5.0)
Not married	305 (95.6)	303 (94.4)	608 (95.0)
Number of children, n (%)			
0	293 (91.9)	286 (89.1)	579 (90.5)
1	16 (5.0)	21 (6.5)	37 (5.8)
≥2	10 (3.1)	14 (4.4)	24 (3.8)
Indigenous origin (ethnicity), n (%)			
Aymara	181 (56.7)	179 (55.8)	360 (56.3)
Guarani	1 (0.3)	3 (0.9)	4 (0.6)
Quechua	13 (4.1)	6 (1.9)	19 (3.0)
Other indigenous	10 (3.1)	10 (3.1)	20 (3.1)
No indigenous origin	114 (35.7)	123 (38.3)	237 (37.0)
Occupation, n (%)			
Studying at school	61 (19.1)	58 (18.0)	119 (18.6)
Studying at university	176 (55.2)	181 (56.4)	357 (55.8)
Working	30 (9.4)	36 (11.2)	66 (10.3)
Training	19 (6.0)	17 (5.3)	36 (5.6)
Not working	4 (1.3)	4 (1.3)	8 (1.3)
Working and studying	29 (9.1)	25 (7.8)	54 (8.4)
Highest level of education completed, n (%)			
Primary	19 (6.0)	13 (4.1)	32 (5.0)
Secondary	227 (71.2)	235 (73.2)	462 (72.2)
University	68 (21.3)	62 (19.3)	130 (20.3)
Technical	5 (1.6)	11 (3.4)	16 (2.5)
Baseline method, n (%)			
None	240 (75.2)	257 (80.1)	497 (77.7)
Male condom	46 (14.4)	36 (11.2)	82 (12.8)
Female condom	9 (2.8)	4 (1.3)	13 (2.0)
Other	24 (7.5)	24 (7.5)	48 (7.5)
At least one effective method is acceptable, n (%)			
Yes	26 (8.2)	26 (8.1)	52 (8.1)
No	293 (91.9)	295 (91.9)	588 (91.9)
Pill acceptability, n (%)			
Yes	2 (0.6)	5 (1.6)	7 (1.1)
No	317 (99.3)	316 (98.4)	633 (98.9)
Intrauterine device acceptability, n (%)			

Characteristic	Control (n=319)	Intervention (n=321)	All participants (N=640)
Yes	6 (1.9)	4 (1.3)	10 (1.6)
No	313 (98.1)	317 (98.8)	630 (98.4)
Injection acceptability, n (%)			
Yes	9 (2.8)	6 (1.9)	15 (2.3)
No	310 (97.2)	315 (98.1)	625 (97.7)
Implant acceptability, n (%)			
Yes	5 (1.6)	10 (3.1)	15 (2.3)
No	314 (98.4)	311 (96.9)	625 (97.7)
Patch acceptability, n (%)			
Yes	11 (3.5)	8 (2.5)	19 (3.0)
No	308 (96.6)	313 (97.5)	621 (97.0)

Loss to Follow-Up

Overall, 67.0% (429/640) participants completed the trial follow-up for the coprimary outcome use (control, n=215 and intervention, n=214; Figure 1), and 63.4% (406/640) participants completed the follow-up for the coprimary outcome acceptability (control, n=203 and intervention, n=203). Retention did not differ between the arms (215/319, 67.4% in the control arm and 214/321, 66.7% in the intervention arm; *P*=.84). Among the participants who completed the use coprimary outcome, the strongest predictor of retention was being aged 20 to 24 years (OR 1.33, 95% CI 0.96 to 1.86; *P*=.09). There was some evidence that the effect of this predictor differed by arm (interaction test *P*=.09). Detailed characteristics of participants who completed follow-up and those who did not are reported in Multimedia Appendix 1.

Primary Outcomes

In the intervention arm, 37.4% (80/214) of participants reported the use of effective contraception compared with 33.5% (72/215) of participants in the control arm (Table 2). There was no evidence of a difference in use between the groups (crude OR 1.19, 95% CI 0.80 to 1.76; *P*=.40 and adjusted OR 1.19, 95% CI 0.80 to 1.77; *P*=.40).

Secondary Outcomes

There were no significant differences in any of the secondary outcomes between the groups (Table 3).

In the intervention arm, 71.9% (146/203) of the participants reported that at least one method of contraception was acceptable compared with 62.6% (127/203) of participants in the control arm (Table 2). There was borderline evidence of a difference in acceptability between the groups (crude OR 1.53, 95% CI 1.01 to 2.33; *P*=.05 and adjusted OR 1.49, 95% CI 0.98 to 2.28; *P*=.06).

Process Outcomes

There were no significant differences in any of the process outcomes between the groups (Table 4).

Table 2. Coprimary outcomes.

Outcomes	Control		Intervention		Adjusted odds ratio (95% CI)	*P* value
	N	n (%)	N	n (%)		
Use of effective contraception[a]	215	72 (33.5)	214	80 (37.4)	1.19 (0.80 to 1.77)	.40
At least one effective method is acceptable[a]	203	127 (62.6)	203	146 (71.9)	1.49 (0.98 to 2.28)	.06

[a] Adjusted for age, number of children, education level, and acceptability at baseline.

Table 3. Secondary outcomes.

Outcomes	Control		Intervention		Adjusted odds ratio (95% CI)	*P* value
	N	n (%)	N	n (%)		
Pill acceptability[a]	206	52 (25.2)	207	59 (28.5)	1.19 (0.76 to 1.85)	.45
Intrauterine device acceptability[a]	206	43 (20.8)	206	55 (26.7)	1.37 (0.86 to 2.19)	.18
Injection acceptability[a]	208	79 (38.0)	207	93 (44.9)	1.30 (0.88 to 1.94)	.19
Implant acceptability[a]	206	63 (30.6)	205	65 (31.7)	1.03 (0.68 to 1.58)	.89
Patch acceptability[a]	208	95 (45.7)	208	109 (52.4)	1.31 (0.89 to 1.93)	.17
Long-acting reversible contraception acceptability[a]	204	106 (52.0)	205	120 (58.5)	1.31 (0.88 to 1.93)	.18
Effective contraceptive use during the 4 months[b]	210	76 (36.2)	206	73 (35.4)	0.94 (0.62 to 1.40)	.76
Service uptake[b] (attended a service one or more times)	210	110 (52.4)	205	93 (45.4)	0.74 (0.50 to 1.10)	.14
Unintended pregnancy	319	1 (0.3)	321	0 (0.0)	N/A[c]	N/A
Induced abortion[d]	209	3 (1.4)	205	1 (0.5)	0.34 (0.01 to 4.24)[d]	.64

[a] Adjusted for age, number of children, education level, and the corresponding method’s acceptability at baseline.

[b] Adjusted for age, number of children, education level, and acceptability at baseline.

[c] N/A: not applicable.

[d] Unadjusted exact logistic regression.

Table 4. Process outcomes.

Process outcome	Control	Intervention	Proportional odds ratio[a] (95% CI)	*P* value
Knowledge of effective contraception, mean (SD)	4.3 (1.9)	4.5 (1.8)	0.17[b] (−0.19 to 0.53)	.36
My friends would use the pill, IUD [c], injection, or implant if they wanted to prevent pregnancy (N=205 for Control; N=202 for Intervention) , n (%)			1.17 (0.73 to 1.88)	.51
Strongly disagree	2 (1.0)	0 (0)		
Disagree	7 (3.4)	2 (1.0)		
Not sure	29 (14.2)	13 (15.6)		
Agree	159 (77.6)	161 (79.7)		
Strongly agree	8 (3.9)	8 (4.0)		
My friends would talk to their partner about contraception if they wanted to prevent a pregnancy (N=205 for Control; N=202 for Intervention) , n (%)			1.33 (0.91 to 1.94)	.15
Strongly disagree	1 (0.5)	0 (0)		
Disagree	17 (8.3)	14 (6.9)		
Not sure	79 (38.5)	67 (33.2)		
Agree	105 (51.2)	118 (58.4)		
Strongly agree	3 (1.5)	3 (1.5)		
If you wanted to use the pill, IUD, injection, or implant, how easy would it be for you to use it? (N=205 for Control; N=202 for Intervention), n (%)			0.98 (0.64 to 1.51)	.93
Very difficult	4 (2.0)	2 (1.0)		
Difficult	24 (11.7)	25 (12.4)		
Not sure	17 (8.3)	17 (8.4)		
Easy	149 (72.7)	149 (73.8)		
Very easy	11 (5.4)	9 (4.5)		
If you wanted to talk to your partner about contraception, how easy would it be for you to talk to him? (N=205 for Control; N=202 for Intervention) , n (%)			0.71 (0.48 to 1.06)	.09
Very difficult	1 (0.5)	7 (3.5)		
Difficult	22 (10.7)	27 (13.4)		
Not sure	33 (16.1)	25 (12.4)		
Easy	128 (62.4)	136 (67.3)		
Very easy	21 (10.2)	7 (3.5)		
If you wanted to use the pill, IUD, injection, or implant, how certain are you that you could use it? (N=205 for Control; N=202 for Intervention), n (%)			1.01 (0.66 to 1.55)	.97
Very certain I could not	2 (1.0)	2 (1.0)		
Certain I could not	2 (1.0)	2 (1.0)		
Not sure	34 (16.6)	36 (17.8)		
Certain I could	151 (73.7)	143 (70.8)		
Very certain I could	16 (7.8)	19 (9.4)		
If you wanted to talk to your partner about contraception, how certain are you that you could talk to him? (N=204 for Control; N=202 for Intervention), n (%)			0.87 (0.58 to 1.30)	.49

Process outcome	Control	Intervention	Proportional odds ratio[a] (95% CI)	*P* value
Very certain I could not	0 (0)	4 (2.0)		
Certain I could not	9 (4.4)	6 (3.0)		
Not sure	46 (22.6)	45 (22.3)		
Certain I could	131 (64.2)	137 (67.8)		
Very certain I could	18 (8.8)	10 (5.0)		
I intend to use the pill, IUD, injection, implant or patch (N=204 for Control; N=202 for Intervention), n (%)			0.74 (0.50 to 1.10)	.14
Strongly disagree	3 (1.5)	2 (1.0)		
Disagree	14 (6.9)	16 (7.9)		
Not sure	18 (8.8)	30 (14.9)		
Agree	134 (65.7)	125 (61.9)		
Strongly agree	35 (17.2)	29 (14.4)		
Number of messages read (N=206), n (%)			N/A[d]	N/A
All	N/A	13 (6.3)		
Most	N/A	40 (19.4)		
Some	N/A	94 (45.6)		
None	N/A	59 (28.6)		
Proportion of intervention participants that stopped the intervention (N=205), n (%)	N/A	23 (11.2)	N/A	N/A

[a] Estimated from ordered logistic regression.

[b] Mean difference.

[c] IUD: intrauterine device.

[d] N/A: not applicable.

Potential for Contamination

A total of 1.0% (2/209) of control participants said that they read the messages of someone else in the study. Moreover, 3.9% (8/205) of intervention participants said that someone else in the study read their messages.

Report of Physical Violence

A total of 2.9% (6/207) of participants in the control group and 2.0% (4/202) in the intervention group reported that they experienced physical violence since being in the study (Fisher exact test *P*=.75).

Intervention Dose

A total of 25.7% (53/206) of intervention participants reported that they read all or most of the intervention messages, with 28.6% (59/206) stating that they read none of the messages. In addition, 11.2% (23/205) of the participants reported that they stopped the intervention messages. Reasons intervention participants provided for not reading the messages or uninstalling the app were concerns about confidentiality, the app took up too much space on their phone, there were too many messages, and some messages were repetitive. In addition, 18.9% (39/206) of the intervention participants who answered the open-ended question “Did anything good or bad happened as a result of receiving the messages?” said that they did not receive any messages.

Participants’ Report of Anything Good or Bad That Happened During the Study

Almost half of the intervention participants that answered this question (97/206) reported something positive about the messages. The most common comment was that they learned new information. One participant said that they got pregnant and another said that they had *a scare due to carelessness.*

Sensitivity Analyses

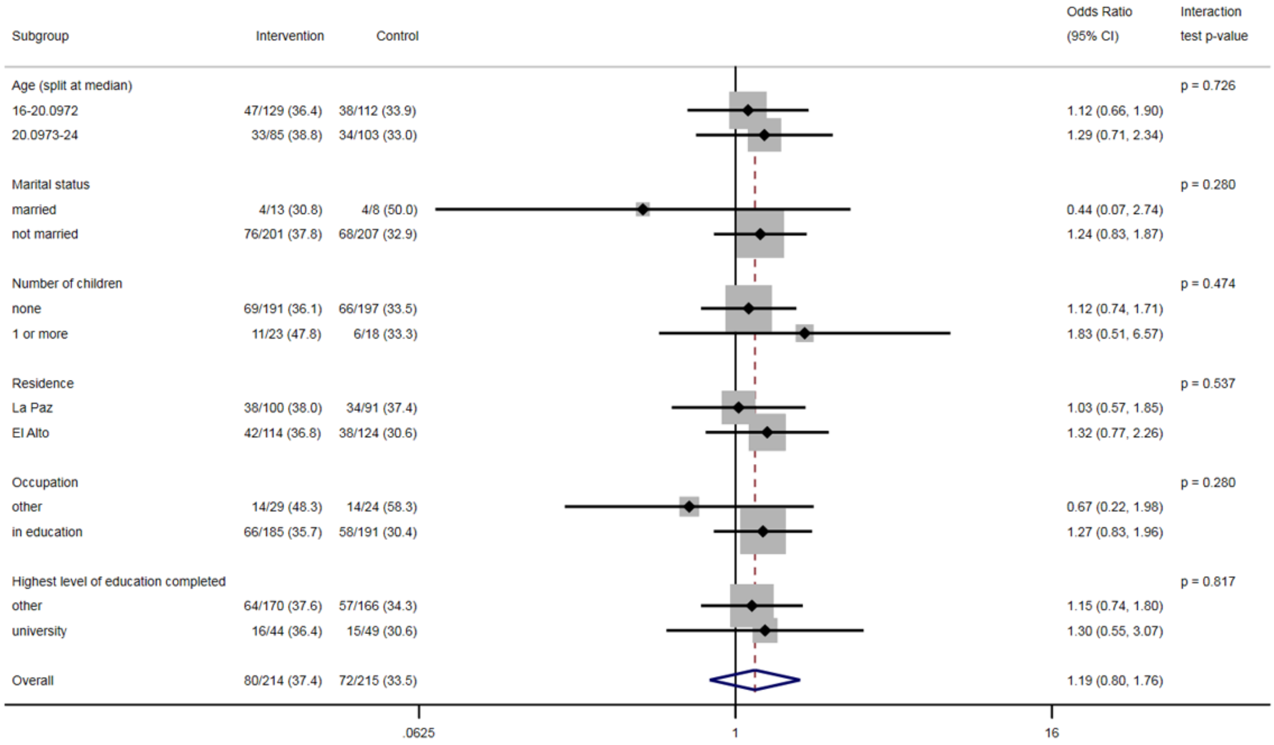
When we considered that participants who were lost to follow-up did not use an effective method or find an effective method acceptable, the effects observed in the principal analysis were reduced (use: OR 1.14, 95% CI 0.79-1.64; *P*=.48 and acceptability: OR 1.26, 95% CI 0.92-1.74; *P*=.15).

The strongest predictor of retention was being aged 20 to 24 years. Age was a baseline covariate, so the model in the second sensitivity analysis (adjusting for the main baseline predictors of missingness) is the same as the primary analysis model.

Subgroup Analysis

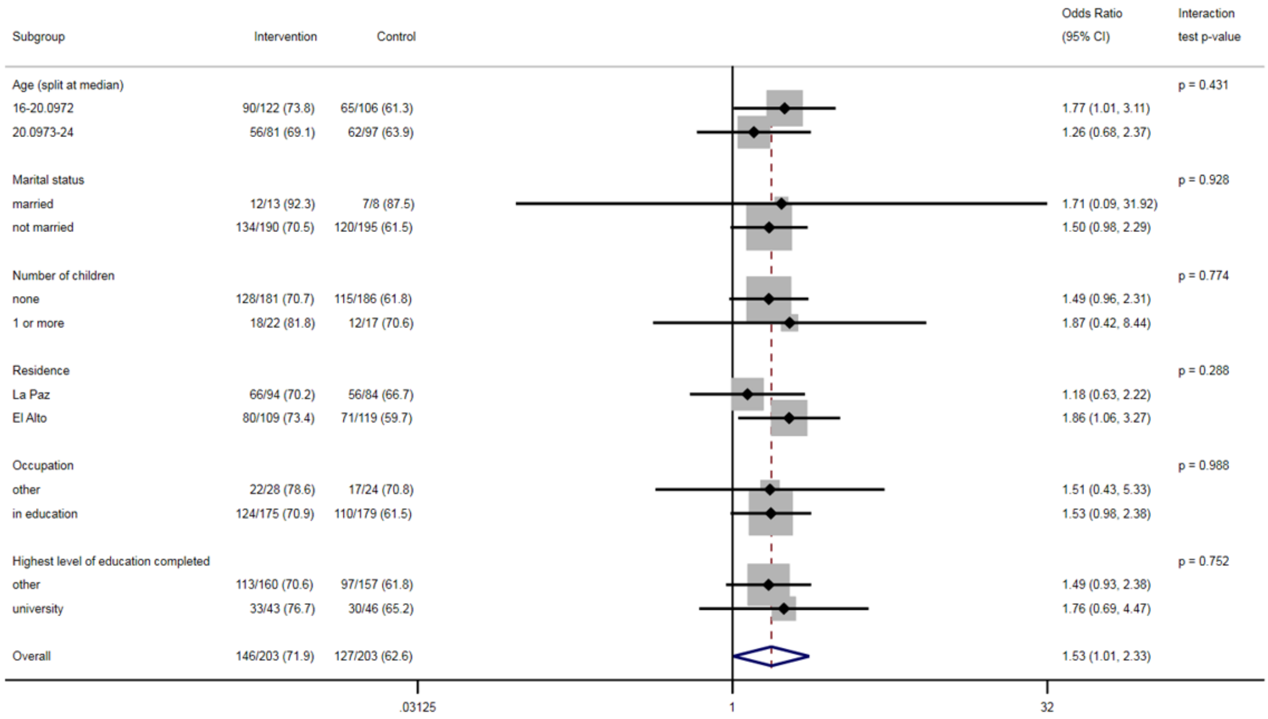
There was no evidence that the effect of the intervention differed within the different levels of the subgroups (Figures 2 and 3).

Figure 2. Intervention effect on the use of effective contraception by subgroups.



Subgroup	Intervention	Control	Odds Ratio (95% CI)	Interaction test p-value
Age (split at median)				p = 0.726
16-20.0972	47/129 (36.4)	38/112 (33.9)	1.12 (0.66, 1.90)	
20.0973-24	33/85 (38.8)	34/103 (33.0)	1.29 (0.71, 2.34)	
Marital status				p = 0.280
married	4/13 (30.8)	4/8 (50.0)	0.44 (0.07, 2.74)	
not married	76/201 (37.8)	68/207 (32.9)	1.24 (0.83, 1.87)	
Number of children				p = 0.474
none	69/191 (36.1)	66/197 (33.5)	1.12 (0.74, 1.71)	
1 or more	11/23 (47.8)	6/18 (33.3)	1.83 (0.51, 6.57)	
Residence				p = 0.537
La Paz	38/100 (38.0)	34/91 (37.4)	1.03 (0.57, 1.85)	
El Alto	42/114 (36.8)	38/124 (30.6)	1.32 (0.77, 2.26)	
Occupation				p = 0.280
other	14/29 (48.3)	14/24 (58.3)	0.67 (0.22, 1.98)	
in education	66/185 (35.7)	58/191 (30.4)	1.27 (0.83, 1.96)	
Highest level of education completed				p = 0.817
other	64/170 (37.6)	57/166 (34.3)	1.15 (0.74, 1.80)	
university	16/44 (36.4)	15/49 (30.6)	1.30 (0.55, 3.07)	
Overall	80/214 (37.4)	72/215 (33.5)	1.19 (0.80, 1.76)	

Figure 3. Intervention effect on the acceptability of effective contraception by subgroups.



Subgroup	Intervention	Control	Odds Ratio (95% CI)	Interaction test p-value
Age (split at median)				p = 0.431
16-20.0972	90/122 (73.8)	65/106 (61.3)	1.77 (1.01, 3.11)	
20.0973-24	56/81 (69.1)	62/97 (63.9)	1.26 (0.68, 2.37)	
Marital status				p = 0.928
married	12/13 (92.3)	7/8 (87.5)	1.71 (0.09, 31.92)	
not married	134/190 (70.5)	120/195 (61.5)	1.50 (0.98, 2.29)	
Number of children				p = 0.774
none	128/181 (70.7)	115/186 (61.8)	1.49 (0.96, 2.31)	
1 or more	18/22 (81.8)	12/17 (70.6)	1.87 (0.42, 8.44)	
Residence				p = 0.288
La Paz	66/94 (70.2)	56/84 (66.7)	1.18 (0.63, 2.22)	
El Alto	80/109 (73.4)	71/119 (59.7)	1.86 (1.06, 3.27)	
Occupation				p = 0.988
other	22/28 (78.6)	17/24 (70.8)	1.51 (0.43, 5.33)	
in education	124/175 (70.9)	110/179 (61.5)	1.53 (0.98, 2.38)	
Highest level of education completed				p = 0.752
other	113/160 (70.6)	97/157 (61.8)	1.49 (0.93, 2.38)	
university	33/43 (76.7)	30/46 (65.2)	1.76 (0.69, 4.47)	
Overall	146/203 (71.9)	127/203 (62.6)	1.53 (1.01, 2.33)	

Discussion

Principal Findings

The observed difference in contraceptive use between the groups (absolute difference=3.89%) was smaller than expected. Although the use was higher in the intervention group, the difference was not statistically significant. There was a borderline significant effect regarding the acceptability coprimary outcome, which favored the intervention group. No statistically significant differences between the groups in any of the secondary or process outcomes were observed. The intervention dose received was low, based on participants' report.

Strengths and Limitations

The main limitations of this trial were that we did not recruit to target and achieved less than 80% follow-up completion.

Although effect estimates or both primary outcomes favored sending the intervention messages with the Tú decides app, the differences between the groups were smaller than what we expected. The recruitment and follow-up challenges meant that the trial was underpowered and therefore unable to produce unequivocal estimates regarding the effect of sending the intervention messages in addition to the app (with 429 participants, the trial had 54% power to detect a 10% absolute difference in use of effective contraception between the groups). Another limitation is in relation to the self-reported outcome measures. As they are self-reported, they are more likely to be biased than if the outcomes were objective, such as clinic-verified use of contraception.

Despite the limitations, this study had several strengths. Our trial database and randomization system generated and concealed the allocated and achieved well-balanced groups. There was no evidence that the intervention was associated with an increase in self-reported violence, a potential adverse outcome related to others viewing the messages, unwanted by the participant. We considered this a potential adverse outcome, given the stigma associated with sex before marriage in Bolivia [36]. However, we cannot determine the effect of the app on partner violence because both groups had access to it. Despite this, it is reassuring that the self-reported prevalence in this trial was low (2%).

Comparisons With Existing Research

Trials evaluating contraceptive behavioral interventions delivered by mobile phone have had mixed results [25-35], with some showing an improvement in contraceptive use [25-27] and knowledge [25,28-30]. The results of this trial are not inconsistent with our trials of similar interventions among young people in Tajikistan [50] and Palestine [51]. In the Palestine trial, participants who received the intervention were more than twice as likely to find at least one method of contraception acceptable (OR 2.34, 95% CI 1.48-3.68; $P<.001$). There were also improvements in knowledge, acceptability of individual methods, perceived norms about friends using contraception, and intention to use contraception compared with the control group. In the Tajikistan trial, there was contamination between the intervention and control groups, and no differences were found between the groups. As with this Bolivian trial, the Tajik and Palestinian trials also did not suggest that the intervention was associated with an increase in violence.

Implications of the Findings

The uncertainty regarding the efficacy of the intervention means that we cannot strongly recommend implementation in Bolivia. The low dose of the intervention is likely to have reduced the effect estimates. Only 25.7% (53/206) of the participants reported that they read all or most of the intervention messages, with 28.6% (59/206) reporting that they read none of the messages. This could have contributed to the smaller than anticipated observed differences. As we did not collect data on why these people did not read most of the messages, we do not know the reasons for this. A total of 47% (97/206) of intervention participants who answered the question “Did anything good or bad happen during the study?” (97/321, 30.2% of all intervention participants and 97/214, 45.3% of those who completed follow-up) reported something positive about the messages. Although participants in the development work were positive about the message content, sending the messages to the target group as they are intended to be sent (eg, over the entire 4 months) and then interviewing participants about their experience may have identified barriers to successful intervention receipt. We have done this in previous studies, but our timeline and resources did not allow for this in this trial.

Another possibility for the smaller than anticipated differences could be that the intervention is only moderately more effective than offering standard family planning information on a mobile app. It may be that offering good-quality family planning information on mobile app pages in this context is sufficient enough to improve the use of and attitudes toward contraception. However, the results suggest that the intervention messages could increase the acceptability of effective contraception if they were offered alongside the download of the Tú decides app and would not cause harm if done so.

It is difficult to determine exactly why we were unable to recruit to target. Some factors that likely contributed to the under-recruitment are (1) the trial promotion was not targeted to eligible people as much as it could have been, (2) young women were reticent to admit to being sexually active, (3) they did not want to have the app on their phone (although the development work indicated that this would not be a problem for them), and (4) they were not interested in taking part in a trial. Although the target sample size was much lower in the Tajik and Palestine trials (n=570 for each), this Bolivian trial actually recruited more participants and had a narrower inclusion criterion (women aged 16-24 years with an unmet need for contraception). Potential ways to improve recruitment in future trials of contraceptive behavioral interventions delivered by a mobile phone could be to promote the trial in settings where recruiters have a very good chance of accessing eligible people. In addition, only assessing potential participants on their own may improve the chance that they would admit to being sexually active.

In deciding how to analyze the scales, we thought it would be better to avoid false positives. For participants to score *acceptable* for a method, they must have chosen *agree* or *strongly agree* to the positively worded stems and *disagree* or *strongly disagree* to the negatively worded stems. To illustrate how stringent this definition of *acceptability* is, a participant could respond *strongly agree* to the statement that the method has unwanted side effects but could still use the method because the benefit of using it (avoiding a pregnancy) outweighs the risk [51]. This does not have implications on the effect of the intervention relative to the control. Nevertheless, the very low baseline acceptability of all methods highlights the need for interventions such as the one evaluated in this trial—interventions that provide accurate, nonjudgmental information about contraception and that address negative beliefs and misconceptions.

Conclusions

This trial was unable to determine unequivocally if the intervention was effective at increasing the use and acceptability of effective contraception among young women with an unmet need in Bolivia. The intervention messages when delivered in addition to an app providing standard family planning information may moderately improve acceptability. Future research could first identify why around one-third of participants did not read the intervention messages and then evaluate the effect of the intervention on the use of contraception in different contexts.

Acknowledgments

The authors would like to thank the young people in Bolivia who participated in the intervention development phase and the trial, promoted the study, and assisted with collection of the follow-up data. The authors are extremely grateful for their hard work and volunteering their time. The authors would also like to thank the International Planned Parenthood Federation for funding the trial and Alison McKinley, the Senior Research and Evaluation Officer for the Innovation Program, for her support throughout the project. The trial was supported by the International Planned Parenthood Federation Innovation Programme.

Authors' Contributions

OM designed and managed the trial, conducted the analysis, and wrote the manuscript. CA coordinated and conducted the trial recruitment and follow-up. MT facilitated trial implementation. JG facilitated trial implementation and took overall local responsibility for the project. SH contributed to planning discussions regarding the trial. IA developed the trial database and randomization system. BL provided advice regarding the statistical analysis and reviewed the Stata analysis code. PE oversaw the statistical analysis. MP managed the trial follow-up when OM was on maternity leave and reviewed the primary outcome Stata analysis code. CF provided guidance regarding the trial design and took overall academic responsibility for the project. All authors revised the work, approved the version to be published, and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics by follow-up status.
[DOCX File , 21 KB - jmir_v22i6e14073_app1.docx]

References

1. Hardee K, Eggleston E, Wong EL, Irwanto, Hull TH. Unintended pregnancy and women's psychological well-being in Indonesia. J Biosoc Sci 2004 Sep;36(5):617-626. [doi: 10.1017/s0021932003006321] [Medline: 15446355]
2. Gipson JD, Koenig MA, Hindin MJ. The effects of unintended pregnancy on infant, child, and parental health: a review of the literature. Stud Fam Plann 2008 Mar;39(1):18-38. [doi: 10.1111/j.1728-4465.2008.00148.x] [Medline: 18540521]
3. Marston C, Cleland J. Do unintended pregnancies carried to term lead to adverse outcomes for mother and child? An assessment in five developing countries. Popul Stud (Camb) 2003;57(1):77-93. [doi: 10.1080/0032472032000061749] [Medline: 12745811]
4. Shah PS, Balkhair T, Ohlsson A, Beyene J, Scott F, Frick C. Intention to become pregnant and low birth weight and preterm birth: a systematic review. Matern Child Health J 2011 Feb;15(2):205-216. [doi: 10.1007/s10995-009-0546-2] [Medline: 20012348]
5. Mohllajee AP, Curtis KM, Morrow B, Marchbanks PA. Pregnancy intention and its relationship to birth and maternal outcomes. Obstet Gynecol 2007 Mar;109(3):678-686. [doi: 10.1097/01.AOG.0000255666.78427.c5] [Medline: 17329520]
6. Sedgh G, Ashford LS, Hussain R. Guttmacher Institute. New York, NY: Guttmacher Institute; 2016 Jun. Unmet Need for Contraception in Developing Countries: Examining Women's Reasons for Not Using a Method URL: https://www.guttmacher.org/sites/default/files/report_pdf/unmet-need-for-contraception-in-developing-countries-report.pdf [accessed 2020-03-13]
7. World Health Organization. 2018 Feb 8. Family Planning/Contraception Fact Sheet URL: http://www.who.int/mediacentre/factsheets/fs351/en/ [accessed 2020-03-13]
8. Cleland J, Ali MM. Reproductive consequences of contraceptive failure in 19 developing countries. Obstet Gynecol 2004 Aug;104(2):314-320. [doi: 10.1097/01.AOG.0000134789.73663.fd] [Medline: 15292005]
9. Trussell J. Contraceptive efficacy. In: Hatcher R, Trussell J, Nelson A, Cates W, Kowal D, Policar M, editors. Contraceptive Technology. New York, NY: Ardent Media; 2011.
10. The DHS Program. 2008. 2008 National Demography and Health Survey. In Spanish. Encuesta Nacional de Demografía y Salud 2008 URL: https://dhsprogram.com/publications/publication-FR228-DHS-Final-Reports.cfm [accessed 2017-10-20]
11. Ministerio de Salud. 2017. Demographia and Health Survey - EDSA 2016. In Spanish. Encuesta de Demographia y Salud - EDSA 2016 URL: https://www.minsalud.gob.bo/images/Documentacion/EDSA-2016.pdf [accessed 2019-03-21] [WebCite Cache ID 772TvjsJn]

12. UNFPA - United Nations Population Fund. 2012 Jan. Independent Country Programme Evaluation-Bolivia 2008-2011 URL: https://www.unfpa.org/sites/default/files/admin-resource/BoliviaReport1_EN_2.pdf [accessed 2017-10-20]
13. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
14. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
15. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
16. Zurovac D, Sudoi RK, Akhwale WS, Ndiritu M, Hamer DH, Rowe AK, et al. The effect of mobile phone text-message reminders on Kenyan health workers' adherence to malaria treatment guidelines: a cluster randomised trial. Lancet 2011 Aug 27;378(9793):795-803 [FREE Full text] [doi: 10.1016/S0140-6736(11)60783-6] [Medline: 21820166]
17. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
18. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
19. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
20. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
21. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
22. Head KJ, Noar SM, Iannarino NT, Harrington NG. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
23. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2011.
24. L'Engle KL, Mangone ER, Parcesepe AM, Agarwal S, Ippoliti NB. Mobile phone interventions for adolescent sexual and reproductive health: a systematic review. Pediatrics 2016 Sep;138(3):pii: e20160884 [FREE Full text] [doi: 10.1542/peds.2016-0884] [Medline: 27553221]
25. Castaño PM, Bynum JY, Andrés R, Lara M, Westhoff C. Effect of daily text messages on oral contraceptive continuation: a randomized controlled trial. Obstet Gynecol 2012 Jan;119(1):14-20. [doi: 10.1097/AOG.0b013e31823d4167] [Medline: 22143257]
26. Harrington EK. ResearchWorks. Evaluation of an mHealth SMS dialogue strategy to meet women’s and couples’ postpartum contraceptive needs in Kenya (Mobile WACh XY): A randomized controlled trial URL: https://digital.lib.washington.edu/researchworks/handle/1773/41691 [accessed 2019-03-21] [WebCite Cache ID 772UEarIf]
27. Smith C, Ngo TD, Gold J, Edwards P, Vannak U, Sokhey L, et al. Effect of a mobile phone-based intervention on post-abortion contraception: a randomized controlled trial in Cambodia. Bull World Health Organ 2015 Dec 1;93(12):842-50A [FREE Full text] [doi: 10.2471/BLT.15.160267] [Medline: 26668436]
28. Hebert LE, Hill BJ, Quinn M, Holl JL, Whitaker AK, Gilliam ML. Mobile contraceptive application use in a clinical setting in addition to standard contraceptive counseling: A randomized controlled trial. Contraception 2018 Oct;98(4):281-287. [doi: 10.1016/j.contraception.2018.07.001] [Medline: 30006030]
29. Johnson D, Juras R, Riley P, Chatterji M, Sloane P, Choi SK, et al. A randomized controlled trial of the impact of a family planning mHealth service on knowledge and use of contraception. Contraception 2017 Jan;95(1):90-97. [doi: 10.1016/j.contraception.2016.07.009] [Medline: 27421767]
30. Rokicki S, Cohen J, Salomon JA, Fink G. Impact of a text-messaging program on adolescent reproductive health: a cluster-randomized trial in Ghana. Am J Public Health 2017 Feb;107(2):298-305. [doi: 10.2105/AJPH.2016.303562] [Medline: 27997236]
31. Bull S, Devine S, Schmiege SJ, Pickard L, Campbell J, Shlay JC. Text messaging, teen outreach program, and sexual health behavior: a cluster randomized trial. Am J Public Health 2016 Sep;106(S1):S117-S124. [doi: 10.2105/AJPH.2016.303363] [Medline: 27689478]
32. Hou MY, Hurwitz S, Kavanagh E, Fortin J, Goldberg AB. Using daily text-message reminders to improve adherence with oral contraceptives: a randomized controlled trial. Obstet Gynecol 2010 Sep;116(3):633-640. [doi: 10.1097/AOG.0b013e3181eb6b0f] [Medline: 20733446]

33. Trent M, Thompson C, Tomaszewski K. Text messaging support for urban adolescents and young adults using injectable contraception: outcomes of the DepoText Pilot Trial. J Adolesc Health 2015 Jul;57(1):100-106 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.03.008] [Medline: 26002432]
34. Tsur L, Kozer E, Berkovitch M. The effect of drug consultation center guidance on contraceptive use among women using isotretinoin: a randomized, controlled study. J Womens Health (Larchmt) 2008 May;17(4):579-584. [doi: 10.1089/jwh.2007.0623] [Medline: 18447762]
35. Sridhar A, Chen A, Forbes ER, Glik D. Mobile application for information on reversible contraception: a randomized controlled trial. Am J Obstet Gynecol 2015 Jun;212(6):774.e1-774.e7. [doi: 10.1016/j.ajog.2015.01.011] [Medline: 25582097]
36. McCarthy OL, Wazwaz O, Calderon VO, Jado I, Saibov S, Stavridis A, et al. Development of an intervention delivered by mobile phone aimed at decreasing unintended pregnancy among young people in three lower middle income countries. BMC Public Health 2018 May 2;18(1):576 [FREE Full text] [doi: 10.1186/s12889-018-5477-7] [Medline: 29716571]
37. Bartholomew EL, Markham C, Ruiter R, Fernandez M, Kok G, Parcel G. Planning Health Promotion Programs: An Intervention Mapping Approach. Fourth Edition. San Francisco, CA: John Wiley & Sons; 2016.
38. Montaño D, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behaviour: Theory, Research and Practice. Fifth Edition. San Francisco, CA: Jossey-Bass; 2015.
39. McCarthy OL, Calderon VO, Makleff S, Huaynoca S, Leurent B, Edwards P, et al. An intervention delivered by app instant messaging to increase acceptability and use of effective contraception among young women in Bolivia: protocol of a randomized controlled trial. JMIR Res Protoc 2017 Dec 18;6(12):e252 [FREE Full text] [doi: 10.2196/resprot.8679] [Medline: 29254910]
40. McCarthy O. Figshare. 2017. A Randomised Controlled Trial of an Intervention Delivered by App Instant Messaging to Increase Acceptability and Use of Effective Contraception Among Young Women in Bolivia: Statistical Analysis Plan URL: https://figshare.com/articles/Bolivia_trial_statistical_analysis_plan/5578012 [accessed 2019-03-21] [WebCite Cache ID 772UTHWod]
41. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
42. Francis J, Eccles M, Johnston M, Walker A, Grimshaw J, Foy R. Constructing Questionnaires Based on the Theory of Planned Behaviour. A Manual for Health Services Researchers. Newcastle upon Tyne, UK: Centre for Health Services Research, University of Newcastle upon Tyne; 2004.
43. Montaño DE, Kasprzyk D, Hamilton DT, Tshimanga M, Gorn G. Evidence-based identification of key beliefs explaining adult male circumcision motivation in Zimbabwe: targets for behavior change messaging. AIDS Behav 2014 May;18(5):885-904 [FREE Full text] [doi: 10.1007/s10461-013-0686-7] [Medline: 24443147]
44. Unidad de Análisis de Políticas Sociales y Económicas. 2014. 2012 Health and Nutrition Assessment Survey: Results Report. In Spanish. Encuesta de evaluación de salud y nutrición 2012: informe de resultados URL: http://www.udape.gob.bo/portales_html/docsociales/LIBROESNUT.pdf [accessed 2020-03-13]
45. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002 Oct 15;21(19):2917-2930. [doi: 10.1002/sim.1296] [Medline: 12325108]
46. Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blümle A, DISCO Study Group. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. Br Med J 2014 Jul 16;349:g4539 [FREE Full text] [doi: 10.1136/bmj.g4539] [Medline: 25030633]
47. Gabler NB, Duan N, Liao D, Elmore JG, Ganiats TG, Kravitz RL. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? Trials 2009 Jun 19;10:43 [FREE Full text] [doi: 10.1186/1745-6215-10-43] [Medline: 19545379]
48. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials 2010 Aug 12;11:85 [FREE Full text] [doi: 10.1186/1745-6215-11-85] [Medline: 20704705]
49. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004 Mar;57(3):229-236. [doi: 10.1016/j.jclinepi.2003.08.009] [Medline: 15066682]
50. McCarthy O, Ahamed I, Kulaeva F, Tokhirov R, Saibov S, Vandewiele M, et al. Correction to: A randomized controlled trial of an intervention delivered by mobile phone app instant messaging to increase the acceptability of effective contraception among young people in Tajikistan. Reprod Health 2018 Mar 26;15(1):52 [FREE Full text] [doi: 10.1186/s12978-018-0496-5] [Medline: 29580246]
51. McCarthy OL, Zghayyer H, Stavridis A, Adada S, Ahamed I, Leurent B, et al. A randomized controlled trial of an intervention delivered by mobile phone text message to increase the acceptability of effective contraception among young women in Palestine. Trials 2019 Apr 23;20(1):228 [FREE Full text] [doi: 10.1186/s13063-019-3297-4] [Medline: 31014358]

Abbreviations

CIES: Centro de Investigación, Educación y Servicios
IBM: Integrated Behavioral Model
IUD: intrauterine device
LSHTM: London School of Hygiene & Tropical Medicine
OR: odds ratio

Edited by G Eysenbach; submitted 21.03.19; peer-reviewed by A Chen, L Brennan, R O'Donnell; comments to author 03.10.19; revised version received 07.11.19; accepted 24.01.20; published 22.06.20.

Please cite as:
McCarthy OL, Aliaga C, Torrico Palacios ME, López Gallardo J, Huaynoca S, Leurent B, Edwards P, Palmer M, Ahamed I, Free C
An Intervention Delivered by Mobile Phone Instant Messaging to Increase Acceptability and Use of Effective Contraception Among Young Women in Bolivia: Randomized Controlled Trial
J Med Internet Res 2020;22(6):e14073
URL: https://www.jmir.org/2020/6/e14073
doi:10.2196/14073
PMID:32568092

Original Paper

Data Validation and Verification Using Blockchain in a Clinical Trial for Breast Cancer: Regulatory Sandbox

Tomonobu Hirano[1], BSc; Tomomitsu Motohashi[1], MSc; Kosuke Okumura[1], BSc; Kentaro Takajo[1]; Taiyo Kuroki[1], MSc; Daisuke Ichikawa[1], MD, PhD; Yutaka Matsuoka[2], MD, PhD; Eisuke Ochi[2,3], PhD; Taro Ueno[1], MD, PhD

[1]SUSMED Inc, Tokyo, Japan
[2]Division of Health Care Research, Center for Public Health Sciences, National Cancer Center Japan, Tokyo, Japan
[3]Faculty of Bioscience and Applied Chemistry, Hosei University, Tokyo, Japan

Corresponding Author:
Taro Ueno, MD, PhD
SUSMED Inc
3-8-5 Honcho, Nihonbashi
Chuo-ku
Tokyo,
Japan
Phone: 81 3 3527 3593
Email: t-ueno@umin.ac.jp

Abstract

Background: The integrity of data in a clinical trial is essential, but the current data management process is too complex and highly labor-intensive. As a result, clinical trials are prone to consuming a lot of budget and time, and there is a risk for human-induced error and data falsification. Blockchain technology has the potential to address some of these challenges.

Objective: The aim of the study was to validate a system that enables the security of medical data in a clinical trial using blockchain technology.

Methods: We have developed a blockchain-based data management system for clinical trials and tested the system through a clinical trial for breast cancer. The project was conducted to demonstrate clinical data management using blockchain technology under the regulatory sandbox enabled by the Japanese Cabinet Office.

Results: We verified and validated the data in the clinical trial using the validation protocol and tested its resilience to data tampering. The robustness of the system was also proven by survival with zero downtime for clinical data registration during a Amazon Web Services disruption event in the Tokyo region on August 23, 2019.

Conclusions: We show that our system can improve clinical trial data management, enhance trust in the clinical research process, and ease regulator burden. The system will contribute to the sustainability of health care services through the optimization of cost for clinical trials.

(J Med Internet Res 2020;22(6):e18938) doi:10.2196/18938

KEYWORDS

blockchain; clinical trial; data management; validation; breast cancer; regulatory sandbox

Introduction

Clinical trials involve a large flow of medical information, and it is necessary to secure the transparency and traceability of clinical data. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use–Good Clinical Practice (ICH-GCP) provides guidance for monitoring the conduct of a clinical trial to verify that reported clinical data are complete, accurate, and accounted for by source records. The long-standing practice in the pharmaceutical and medical device industries has been frequent site visits and 100% source data verification (SDV) on site to ensure that data captured in source records are transferred correctly to the case report forms. It has been estimated that clinical trial monitoring by SDV can consume one-quarter of the sponsor's entire budget for a clinical trial [1]. As the complexity and size of clinical trials grow, it has become increasingly expensive to apply the 100% SDV approach [2,3]. For regulators of trials, such as the US Food and Drug Administration (FDA), European Medicines Agency (EMA) in the European Union, and Pharmaceuticals and Medical Devices Agency (PMDA) in Japan, data auditing is challenging since there is no easy and secure way to access or view complex data. As such, process improvement is an active area of research [4]. With the involvement of more parties and more exchanges, the risk for human-induced error, whether unintentional or malicious, will increase. Previous studies have shown that 17% of authors of clinical trials reported they were personally aware of intentional fabrication in research [5,6]. In the case of misconduct in a clinical trial by an employee of Novartis Pharma, clinical data were manipulated to attribute qualities to the hypertension drug valsartan that it did not possess, such as preventing stroke [7]. The fabricated clinical data were used in an advertising campaign for the drug, and patients were prescribed the drug based on incorrect information. Following misconduct in clinical trials, the Japanese government established and enforced the Clinical Trials Act in 2018 [8]. It obligates clinical researchers to monitor and assure quality, observe practice criteria, and manage conflicts of interest.

Blockchain is emerging as a groundbreaking technology for secure data control in different areas. Bitcoin was the first application of blockchain as a digital currency in extensive use [9]. Recently, researchers have started to focus on using blockchain methodology for building a cryptographic proof of medical systems [10]. They have applied blockchain technology in various health care systems, with potential applications in supply chain management of health care products [11-13], insurance claims processing [14], management of electronic health records and electronic medical records [15-18], maintenance of protocols in clinical trials [19-21], and data management in clinical trials [22-25]. Although there can be many applications of blockchain technology, it is necessary that blockchain be an appropriate technical solution to a particular problem. If there are no incentives for data tampering and all writers can be trusted, blockchain technology is not necessary. On the other hand, if there are incentives for data tampering and it costs a lot to use a trusted third party, such as the Clinical Research Organization, using blockchain makes sense [26]. Due to the frequent occurrences of misconduct in clinical trials and the large consumptions of clinical trial budgets and time by current SDV practices, blockchain is an appropriate technical application in the data management of clinical trials to solve these problems.

We propose a solution to challenges in the current clinical trial system by using blockchain technology coupled with technologies such as client hashchain, encryption protocol, and health check function in servers for hazard management. The system was used in a clinical trial at the National Cancer Center of Japan that investigated the effect of home-based high-intensity interval training for breast cancer survivors. The project was conducted to demonstrate clinical data management using blockchain technology under the regulatory sandbox of the Japanese Cabinet Office [27,28]. This sandbox allows organizations to apply for demonstration and evaluation of new technology, such as blockchain and Internet of Things, without being subject to existing regulations. It simultaneously opens up the possibility for future deregulation measures. During the trial, we experienced the disruption of Amazon Web Services (AWS) cloud servers in the Tokyo region on August 23, 2019 [29]. We report the effect of the blockchain network and health checkup function, which resulted in system survival with zero downtime and secure clinical data registration during the hazardous server shutdown event.

Methods

Breast Cancer Clinical Trial

A clinical trial that investigates the effect of intervention by home-based high-intensity interval training for breast cancer patients was conducted. The study was a parallel-group, single-blind randomized controlled trial. Patients were randomly assigned to the active group with the habit-B program (high-intensity interval training, exercise counseling and guidance, home-based exercise support using information and communication technology, and a wearable device) or treatment as usual with a wearable device.

The eligibility criteria for participants were (1) female, aged between 20 and 59 years at diagnosis; (2) diagnosed with stages I to IIa breast cancer and currently 2 to 13 months postsurgical treatment; (3) not requiring cancer chemotherapy aside from hormone therapy; (4) ability to read, write, and understand Japanese; (5) ability to complete an electronic Patient Reported Outcome Questionnaire via smartphone; (6) consent to trial participation obtained in writing from the patient themselves; and (7) currently engaging in not more than moderate intensity exercise for 30 minutes on two separate days per week (total of 60 minutes).

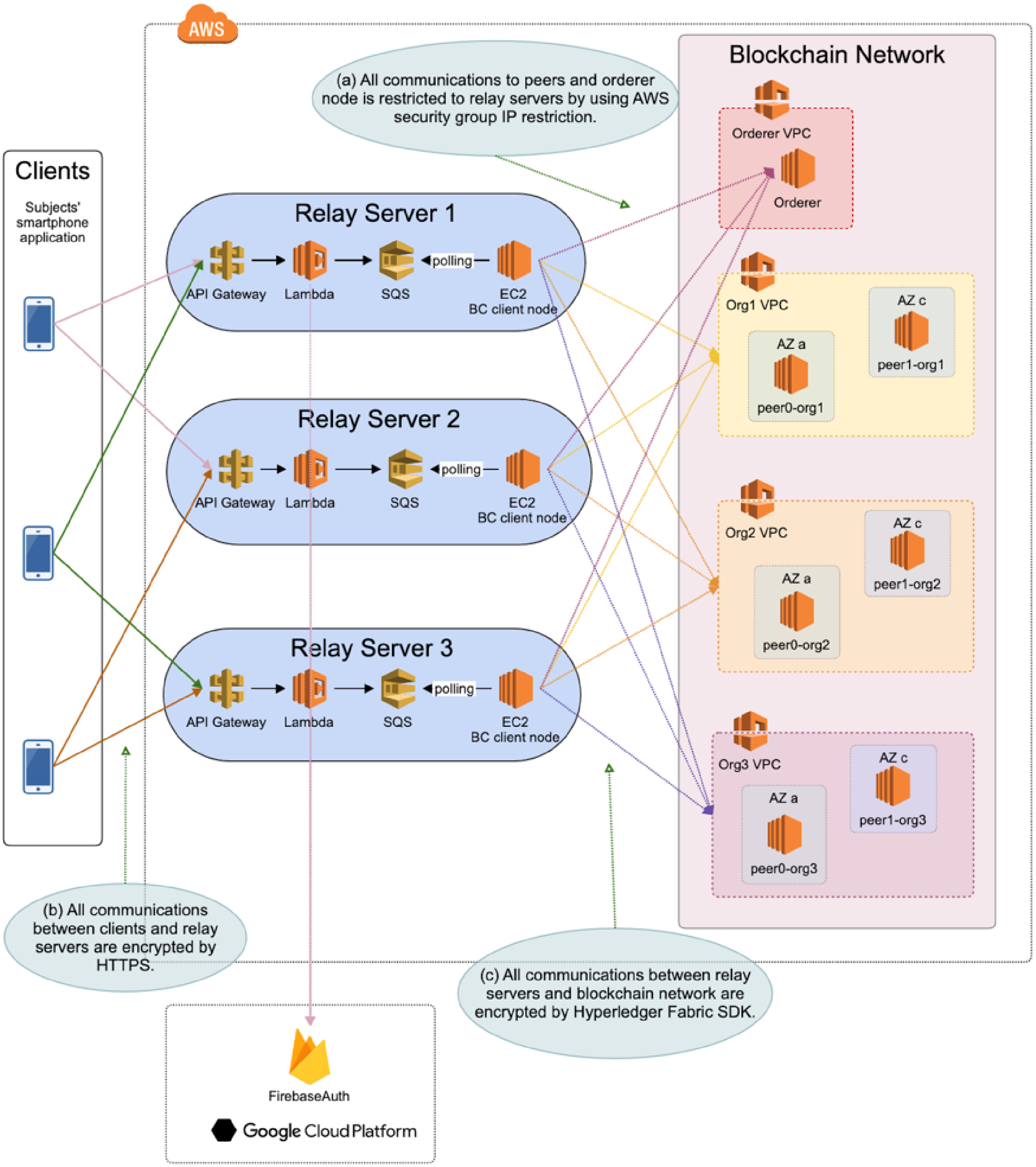
The researcher provided computer-generated random allocation. Using the generated account, the contents of the app that participants used during the clinical trial were assigned automatically to either the habit-B program or the control. When a participant first visits the hospital, she is randomly assigned to an intervention, either with the habit-B program or treatment as usual with a wearable device, after signing informed consent. An independent data center provides computer-generated random allocation as a log-in account for the app. Based on the allocation sequences, the contents of the app participants use during the trial are assigned automatically to either the habit-B program or the control.

The study received ethical approval from the institutional review board of the National Cancer Center Japan and was registered at University Hospital Medical Information Network–Clinical Trial Registry [UMIN000036400]. Participants provided written informed consent to take part in the study. The study began in May 2019 and is ongoing [30].

Architecture of the Data Management System

The data management system that collects the clinical trial data consists of client smartphones, relay servers, and a blockchain network (Figure 1). The relay servers and blockchain network are built on AWS in the Tokyo region. We use Hyperledger Fabric v1.0 (Linux Foundation), an open-source blockchain platform that has been widely used, to operate the blockchain network. Hyperledger Fabric enables throughput of more than 3500 transactions per second in popular deployment configurations [31].

Figure 1. Data management system including client smartphones, relay servers, and blockchain network. The collected data from patient smartphones were sent to the blockchain network via relay servers. We used three relay servers, and the app randomly selected two relay servers to send the data after the authentication of the client device. We have configured the blockchain network to span multiple availability zones in Amazon Web Services.



Data collected from patient smartphones are sent to the blockchain network via relay servers. We use three relay servers, and the app randomly selects two relay servers to send the data after the authentication of the client device. The client app contains an authentication key, and the relay servers verify the key so that the relay server only accepts access from the app for the clinical trial. The relay servers also verify the account for the app using Firebase authentication. These communications between clients and relay servers are encrypted using https. By deploying the relay server and setting the blockchain software development kits to append-only mode, the relay servers send the received data to the blockchain network. The data sent from the relay servers to the blockchain network are encrypted using the authentication protocol by the Hyperledger Fabric. By configuring the internet protocol address restriction to the listed relay proxy, the blockchain network, which contains the clinical trial data, is protected against malicious attack from the external network. The blockchain network is made up of three organizations, each of which contains two validating peers. Each identification and password for relay servers, blockchain network organizations, and nodes of the blockchain network is managed by two independent departments in SUSMED Inc and the National Cancer Center of Japan [22].

We have configured the blockchain network to span multiple availability zones (AZs) in AWS. Each AZ is made up of one or more data centers equipped with independent power, cooling, and networking to ensure fault tolerance. We have also implemented a periodic health check function using the Amazon EC2 Auto Scaling Service. The function continues to maintain a fixed number of instances even if an instance becomes unhealthy. If an instance becomes unhealthy, the group terminates the unhealthy instance and launches another instance to replace it.

In order to increase system durability and prevent data loss, we have configured two kinds of queues in our system. The mobile app puts clinical data into the queue before sending, dequeues and sends the message, and then removes the data from the queue. If relay servers are unavailable, the sending process fails and the data remain available in the queue until the relay servers process it. The other queue is in a relay server system, and the queue is configured using AWS Simple Queue Service (SQS). When a relay server system receives the clinical data, the data are enqueued. In the relay server system, an EC2 instance is responsible for dequeuing and sending the message to the blockchain network. The data are removed from the queue after being successfully processed by the blockchain network. Otherwise, the data remain available in the queue until the sending process succeeds or the SQS retention period expires. In this system, we set the SQS retention period to 7 days. In doing so, we have seven days’ time to recover the blockchain network if there is a breakdown.

Data Collection Using Smartphone

The Global Physical Activity Questionnaire is an internationally standardized questionnaire for surveying physical activity level [32-34]. Fear of cancer recurrence is assessed by the overall fear index score on the Concerns About Recurrence scale [35,36]. Depression is assessed using the Patient Health Questionnaire–9 [37]. Fatigue is assessed by the Cancer Fatigue Scale. Sleep is assessed by the Athens Insomnia Scale [38]. Quality of life is assessed using the EuroQol 5-Dimensions questionnaire [39,40]. All data listed above are stored in JavaScript Object Notation format in the database.

Data Registration to the Blockchain Network

When a participant first logs in to the system, the app generates a secret hash key and preserves it in the client device until the end of the study. The client device calculates a hash value based on the data generated by the patient and the secret hash key as well as the previous hash value using the SHA-256 hash algorithm [22,41]. Thus, the hash value in the client device comprises the chain structure. The hash value is also registered in the blockchain network along with the clinical data in order to guarantee tamper resistance of the value, although the secret hash key was preserved in the client device until the end of the study. The collected data and client hash value are sent to the blockchain network via relay servers. We use three relay servers, and the app randomly selects two relay servers to send the data after the authentication of the client device. Client nodes of the blockchain network are settled in the relay servers and the data are sent from the relay server to the blockchain network. The transactions are validated and accepted by the blockchain network in the following processes:

- Proposal: transaction is sent from client app to the endorsers in each organization
- Endorse: each endorser verifies that (1) the transaction proposal is well formed, (2) it has not already been submitted, (3) the signature is valid, and (4) the client is properly authorized to perform the proposed operation, which is described in the chaincode. If the transaction is validated, the chaincode is executed and the result with the signature is returned to the client
- Submit: client verifies that the number of signatures from the organizations satisfies the endorsement policy. If it is satisfied, the transaction is sent to the ordering service, which orders the series of transactions in chronological order and creates the block of transactions
- Broadcast: the block is delivered to all nodes
- Commit: if each block is well formed and validated to fulfill the endorsement policy, the block is appended to the chain in each node [22]

Data Validation and Verification

When participants complete the data input for the clinical trial, it is necessary to verify all data records sent to the blockchain network to satisfy the following conditions:

- Blockchain network has received all required data types for each participant
- All data records for each participant are correct as a client hashchain
- Blockchain network has received clinical records from two relay servers for each record, and the records are the same

The first condition means that the blockchain network has received all records that are expected to be sent from clients. There are three types of records in the clinical trial. To begin with, a client sends a record that indicates that a secret hash key has been generated. When clinical data are created, the data are sent to the blockchain network via two relay servers. Finally, when the participants complete the clinical trial, the client sends the secret hash key to the blockchain network.

The second condition verifies that all data records registered to the blockchain network are correctly linked as a client hashchain. We sort all data records by their generated time and make a hash value from the clinical data, the secret hash key preserved in the client device until the end of the study, and the previous hash value. Then, we compare the results with the hash value that has been registered to the blockchain network. Due to the sensitivity of a hash function output in relation to its input, changing the data will result in a completely different hash string. Therefore, we can confirm these hash values are generated by the client device that preserves the secret hash key. By verifying this condition, we can detect if a malicious attacker has falsified the clinical data using hash values and clinical data that were stolen.

The purpose of the last condition is to confirm that the relay servers have not been hacked. Clients send their clinical data to two relay servers that are randomly selected from the three relay servers. If malicious users hack a relay server and they send some data from the relay server, the blockchain network will store the data. In order to detect the fraud access and verify the data, we need to confirm that all clinical data is received from two relay servers simultaneously and the data records are identical.

Data Availability

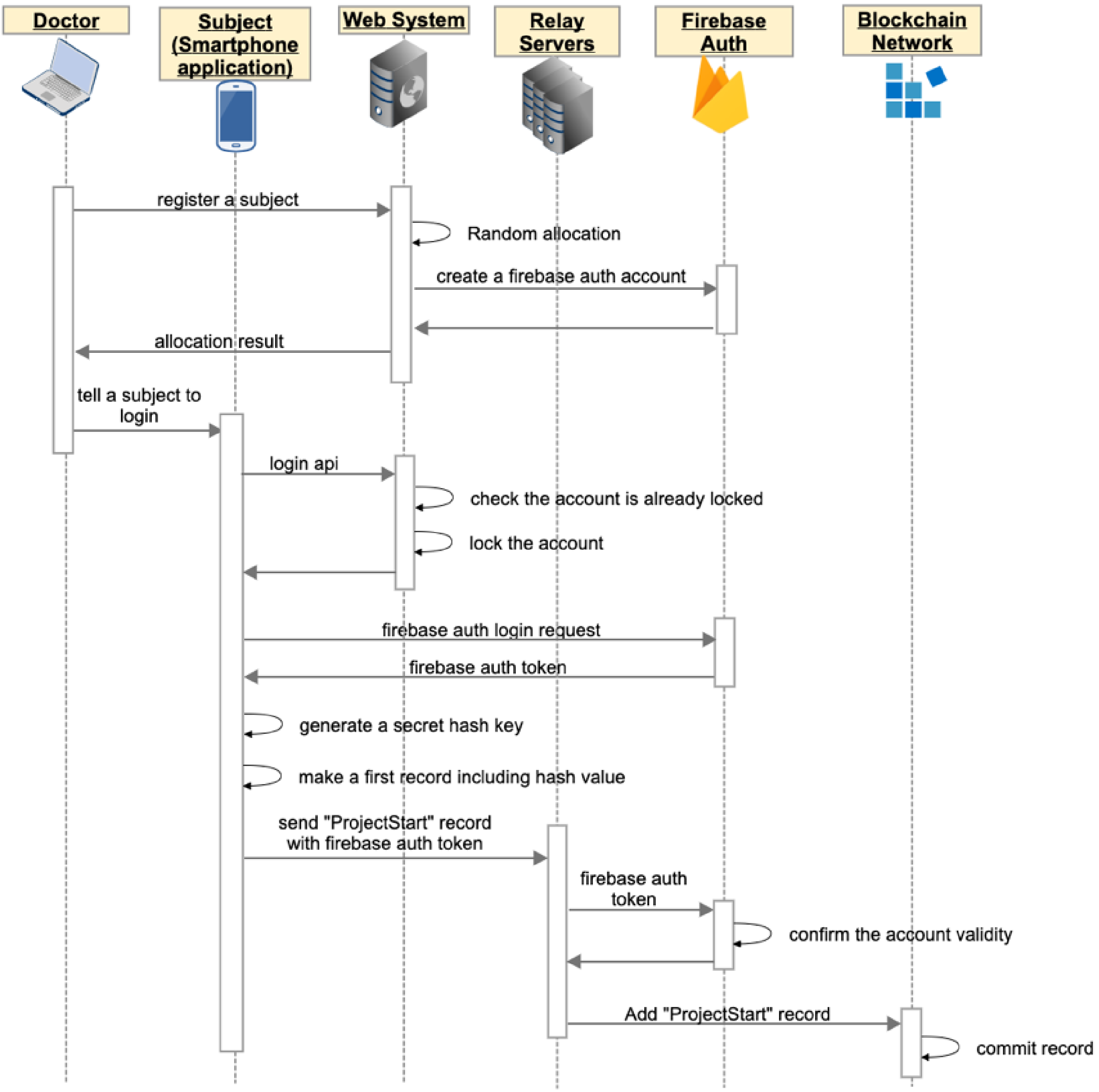
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Results

Registration and Assignment

The participant downloads the app to their own smartphone through the internet and logs in to the system using the account provided by the researcher. When the participant first logs in to the system, the app generates a secret hash key and preserves it in the client's device until the end of the study. The account is locked when the participant first logs in to the system in order to prevent impersonation by stolen account or brute force attack (Figure 2).

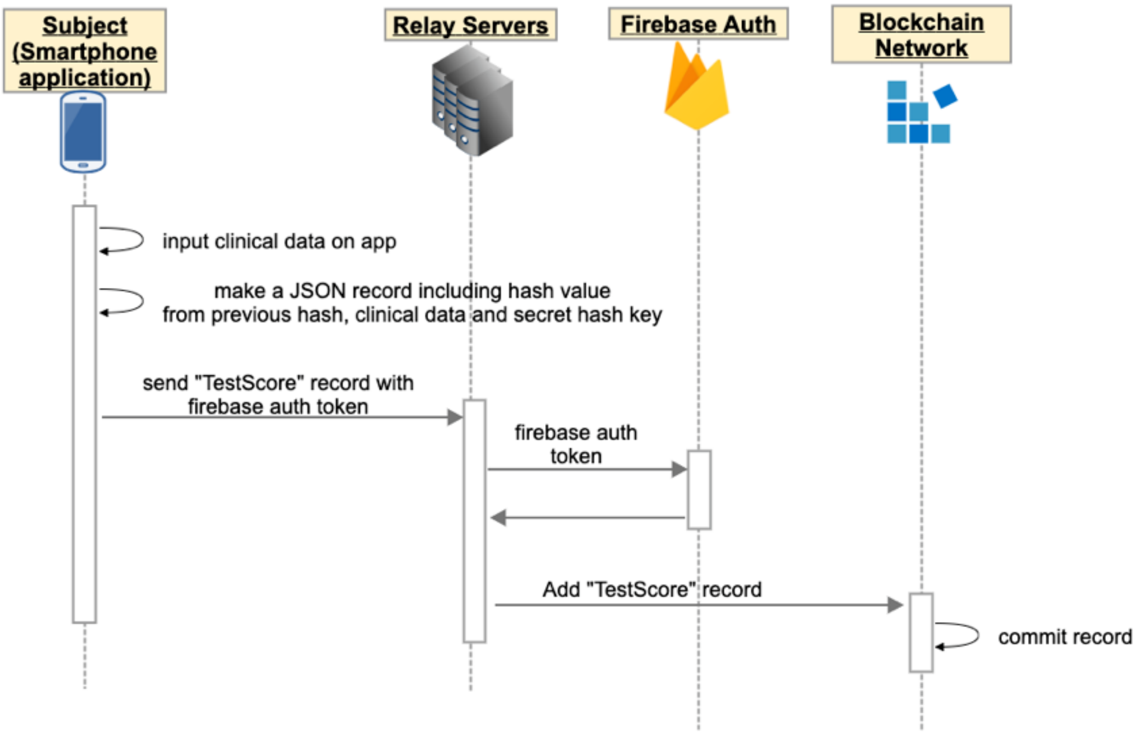
Figure 2. When a participant first visits the hospital, a medical doctor decides if they fit the eligibility criteria. After obtaining informed consent, the doctor registers the patient to the clinical trial using a web-based allocation system. The system automatically creates an account for the participant, and the doctor gives the account to the participant. The participant downloads the app to their smartphone, and they can log in to the app using the account created by the doctor. When the participant first logs in to the app, the account is locked to prevent impersonation, and the secret hash key is then generated. The secret hash key will be preserved in the client app until the participant completes the trial. Using the secret hash key in the client device, the app calculates a hash value. The app sends the first record to the blockchain network via relay servers, and the data is registered in the tamper-resistant blockchain system.



Data Collection

During the trial, the participant inputs the data using the app, and the data are registered to the blockchain network via relay servers (Figure 3). The secret hash key is preserved only in the client’s device during the clinical trial. The client app calculates a hash value based on the medical data, the secret hash key, and the previous hash value using the SHA-256 hash algorithm. The client hash value is also sent to the blockchain network along with the clinical data.

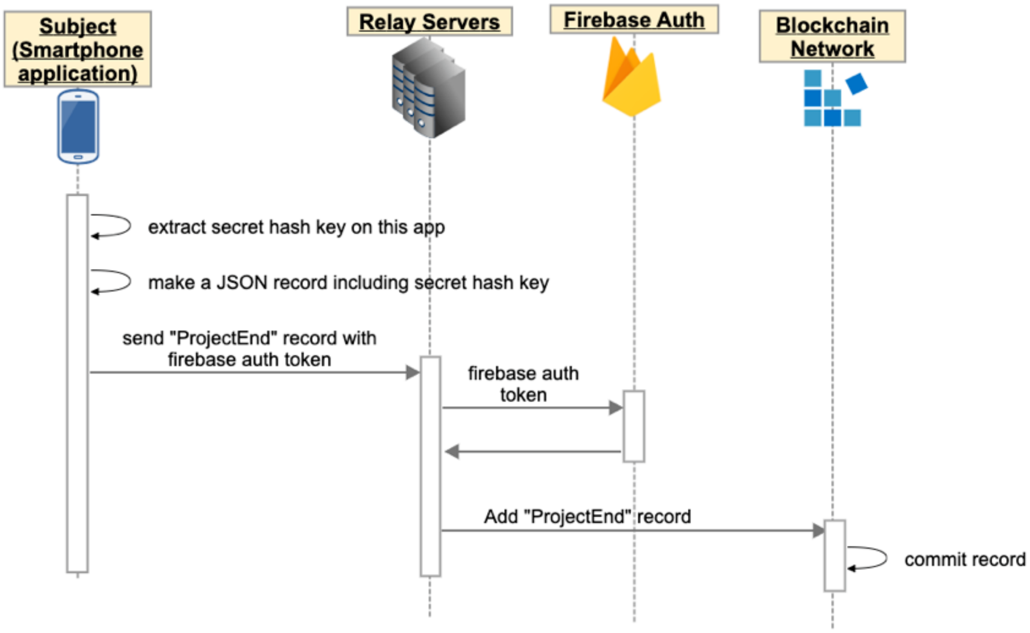
Figure 3. During the clinical trial, participant inputs clinical data using the mobile app. The app calculates the hash value based on the clinical data, the secret hash key preserved in the smartphone until the end of the study, and the previous hash value. The app sends the clinical data and a hash value to the blockchain network via relay servers.



Completion of the Clinical Trial

When the participant completes the clinical trial, the app sends the secret hash key to the blockchain network, and all data of the participant is registered to the blockchain network (Figure 4).

Figure 4. At the end of the clinical trial, the app extracts the secret hash key from the smartphone and sends it to the blockchain network via relay servers.

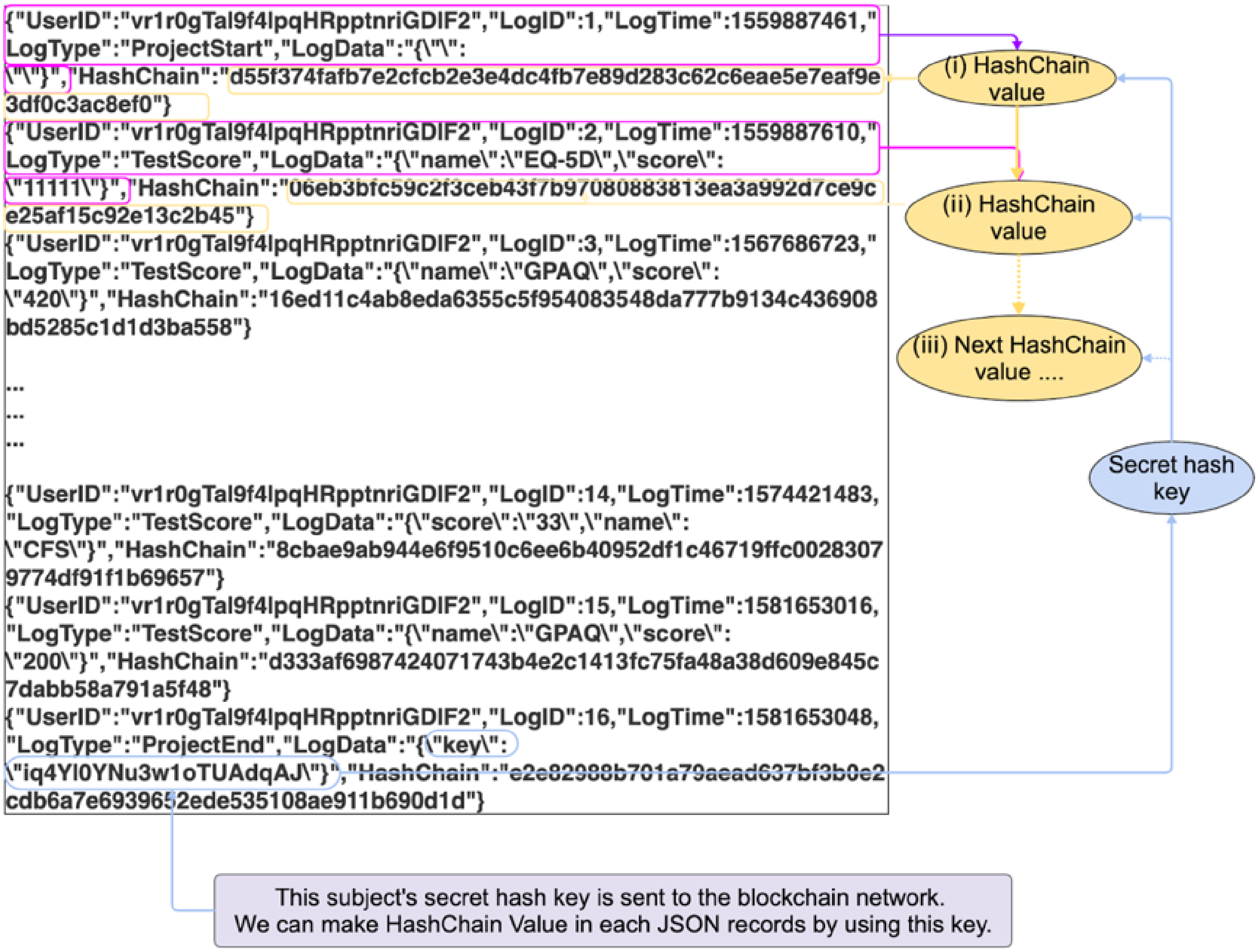


Data Validation and Verification

We verify and validate the patient's data collected in the clinical trial using a predetermined protocol. For regulators such as FDA, EMA and PMDA, the full transaction history since the beginning of the study is readily available with precise log data, and the auditing process can be done quickly and with the reliance that all clinical data are guaranteed or version controlled [23].

Blockchain technology ensures a secure and tamper-proof transaction history so that the blockchain network provides an integrity-protected data storage and process transparency [24]. However, there can be vulnerability before registration to the database in the blockchain network. The data can be tampered with and impersonation could occur, resulting in impairment of reliability of the data. To prevent impersonation of the client, we have configured the mechanism so that the account is locked when the participant first logs in to the system. The client app contains the authentication key, and the relay server checks the key so that the relay server only accepts access from the app for the clinical trial. By configuring the internet protocol address restriction to the listed relay server, the blockchain network, which contains the medical data, is protected against external attack. This configuration makes our system highly secure. Even if an attacker breaks the above security system under impersonation or unauthorized access to the relay server, we can validate and distinguish the correct data using the client hashchain function. In order to complement the tamper resistance of the blockchain network and ensure reliability of the entire system, we verified and validated the data integrity using client hashchains and multiple relay servers. User data along with the client hashchain are registered to the blockchain network. The app sends the user data and client hashchain to the blockchain network during the clinical trial. At the end of the trial, the app sends the secret hash key preserved in the client device to the blockchain network. Using the secret hash key, we verify the integrity of the data by calculating the hashchain retrospectively (Figure 5). User data and the client hashchain are sent from multiple relay servers to the blockchain network. A client app sends its clinical data to two relay servers, which are randomly selected from the three relay servers. In order to detect fraudulent access and verify the data sent from the relay servers, the data are compared with each other to confirm that the records are the same (Figure 6).

Figure 5. Data verification and validation using client hashchain. The first hash value is made from the first data and secret hash key. Subsequent hash values are made from clinical data, the previous hash key, and secret hash key.



```
{"UserID":"vr1r0gTal9f4lpqHRpptnriGDIF2","LogID":1,"LogTime":1559887461,"
LogType":"ProjectStart","LogData":"{\"\":
\"\"}","HashChain":"d55f374fafb7e2cfcb2e3e4dc4fb7e89d283c62c6eae5e7eaf9e
3df0c3ac8ef0"}
{"UserID":"vr1r0gTal9f4lpqHRpptnriGDIF2","LogID":2,"LogTime":1559887610,"
LogType":"TestScore","LogData":"{\"name\":\"EQ-5D\",\"score\":
\"11111\"}","HashChain":"06eb3bfc59c2f3ceb43f7b97080883813ea3a992d7ce9c
e25af15c92e13c2b45"}
{"UserID":"vr1r0gTal9f4lpqHRpptnriGDIF2","LogID":3,"LogTime":1567686723,"
LogType":"TestScore","LogData":"{\"name\":\"GPAQ\",\"score\":
\"420\"}","HashChain":"16ed11c4ab8eda6355c5f954083548da777b9134c436908
bd5285c1d1d3ba558"}


...
...
...

{"UserID":"vr1r0gTal9f4lpqHRpptnriGDIF2","LogID":14,"LogTime":1574421483,
"LogType":"TestScore","LogData":"{\"score\":\"33\",\"name\":
\"CFS\"}","HashChain":"8cbae9ab944e6f9510c6ee6b40952df1c46719ffc0028307
9774df91f1b69657"}
{"UserID":"vr1r0gTal9f4lpqHRpptnriGDIF2","LogID":15,"LogTime":1581653016,
"LogType":"TestScore","LogData":"{\"name\":\"GPAQ\",\"score\":
\"200\"}","HashChain":"d333af6987424071743b4e2c1413fc75fa48a38d609e845c
7dabb58a791a5f48"}
{"UserID":"vr1r0gTal9f4lpqHRpptnriGDIF2","LogID":16,"LogTime":1581653048,
"LogType":"ProjectEnd","LogData":"{\"key\":
\"iq4Yl0YNu3w1oTUAdqAJ\"}","HashChain":"e2e82988b701a79aead637bf3b0e2
cdb6a7e6939652ede535108ae911b690d1d"}
```

This subject's secret hash key is sent to the blockchain network.
We can make HashChain Value in each JSON records by using this key.

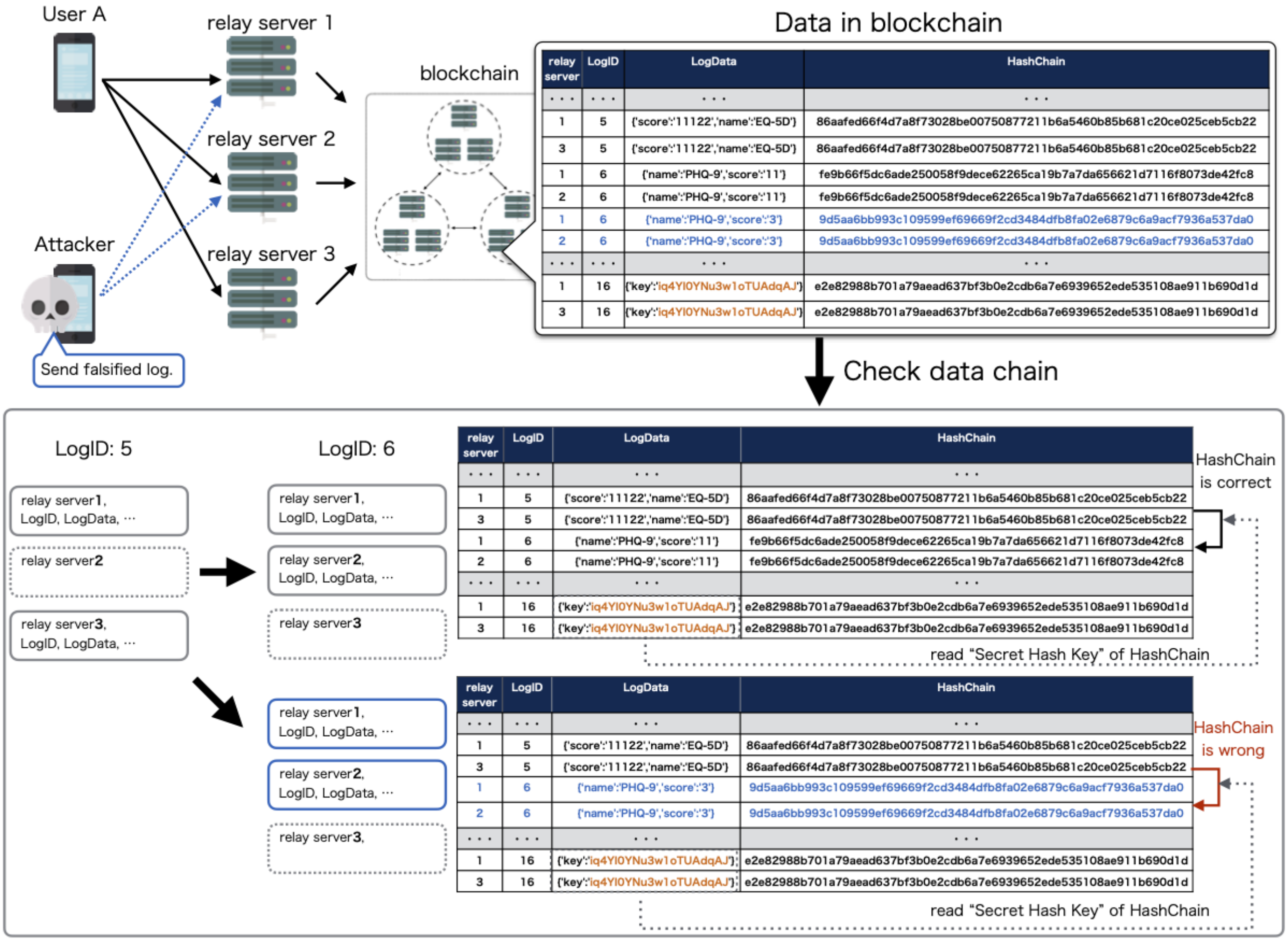
Figure 6. Data verification and validation using records from relay servers.

```
Admin@org1.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","L
ogID":11,"LogTime":1578085488,"LogType":"TestScore","LogData":"{\"name\":
\"GPAQ\",\"score\":
\"420\"}","HashChain":"bb96882e1913b4e5ba24b67042e086be5e8f7dd5b6077721
00db52a01769b1e2"}
Admin@org1.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","L
ogID":12,"LogTime":1578085549,"LogType":"TestScore","LogData":"{\"name\":
\"AIS\",\"score\":
\"2\"}","HashChain":"21fe640b64b9b6d9882e865560f5dc1e552f756cb7b985a49fb8
d904fbedd511"}
...

Admin@org2.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","L
ogID":1,"LogTime":1563516422,"LogType":"ProjectStart","LogData":"","HashChain":
"19c8855619c870ec96794303485388a4421072c4740add386e9734f962b4fc0e"}
Admin@org2.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","L
ogID":10,"LogTime":1578085307,"LogType":"TestScore","LogData":"{\"name\":
\"CFS\",\"score\":
\"17\"}","HashChain":"622963b3687e6860f1cc3b7a792b59eab5400109fd973f972d0
5922c23751234"}
...

Admin@org3.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","L
ogID":1,"LogTime":1563516422,"LogType":"ProjectStart","LogData":"","HashChain":
"19c8855619c870ec96794303485388a4421072c4740add386e9734f962b4fc0e"}
Admin@org3.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","L
ogID":10,"LogTime":1578085307,"LogType":"TestScore","LogData":"{\"name\":
\"CFS\",\"score\":
\"17\"}","HashChain":"622963b3687e6860f1cc3b7a792b59eab5400109fd973f972d0
5922c23751234"}
...
```

Every record starts with Relay Server ID like "Admin@org1.eprobc.susmed.io". We can confirm that every record is sent from at least 2 relay servers and the data records are the same.

To evaluate the validation, we have simulated and tested client impersonation and data falsification by fraudulent access to the relay server. The integrity of the data can be maintained even if attacker tries to manipulate the data by impersonating the client device (Figure 7) or fraudulently accessing the relay servers (Figure 8). The result suggests that our proposed function of client hashchain and multiple relay servers can be a complement to the tamper resistance of the blockchain network to ensure the reliability of the entire system.

Figure 7. Simulation of data falsification by client impersonation. If an attacker successfully logs into the app using a stolen account or brute force attack, the attacker can create clinical data from another client device. The data from the impersonated client will be denied since only the true participant's app preserves the secret hash key that is required to calculate the hashchain.



Data in blockchain

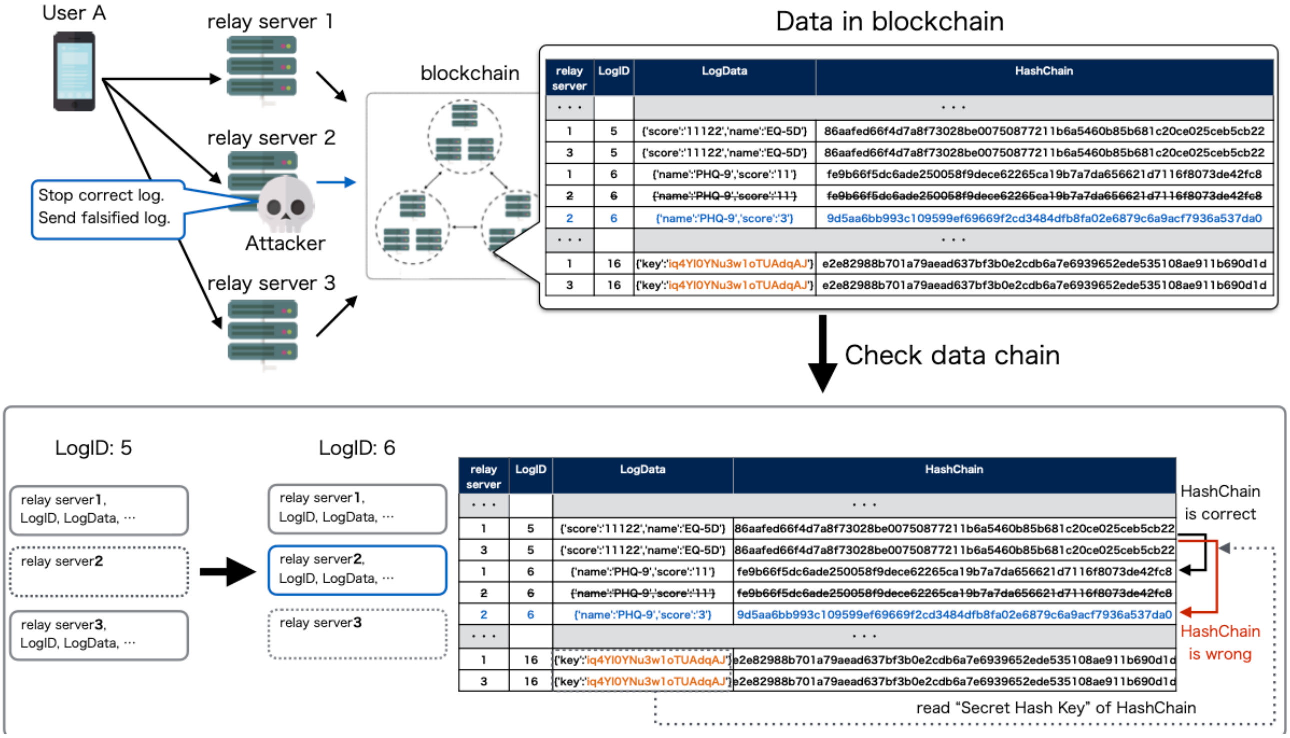
relay server	LogID	LogData	HashChain
· · ·	· · ·	· · ·	· · ·
1	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
3	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
1	6	{'name':'PHQ-9','score':'11'}	fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8
2	6	{'name':'PHQ-9','score':'11'}	fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8
1	6	{'name':'PHQ-9','score':'3'}	9d5aa6bb993c109599ef69669f2cd3484dfb8fa02e6879c6a9acf7936a537da0
2	6	{'name':'PHQ-9','score':'3'}	9d5aa6bb993c109599ef69669f2cd3484dfb8fa02e6879c6a9acf7936a537da0
· · ·	· · ·	· · ·	· · ·
1	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d
3	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d

Check data chain

relay server	LogID	LogData	HashChain
· · ·	· · ·	· · ·	· · ·
1	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
3	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
1	6	{'name':'PHQ-9','score':'11'}	fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8
2	6	{'name':'PHQ-9','score':'11'}	fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8
· · ·	· · ·	· · ·	· · ·
1	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d
3	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d

relay server	LogID	LogData	HashChain
· · ·	· · ·	· · ·	· · ·
1	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
3	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
1	6	{'name':'PHQ-9','score':'3'}	9d5aa6bb993c109599ef69669f2cd3484dfb8fa02e6879c6a9acf7936a537da0
2	6	{'name':'PHQ-9','score':'3'}	9d5aa6bb993c109599ef69669f2cd3484dfb8fa02e6879c6a9acf7936a537da0
· · ·	· · ·	· · ·	· · ·
1	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d
3	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d

Figure 8. Simulation of data falsification by fraudulent access to the relay server. If an attacker successfully accesses the relay server fraudulently, the attacker can send falsified data to the blockchain. The falsified data from the hacked server will be detected by the integrity of the hashchain.



Data in blockchain

relay server	LogID	LogData	HashChain
· · ·		· · ·	
1	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
3	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
1	6	{'name':'PHQ-9','score':'11'}	fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8
2	6	~~{'name':'PHQ-9','score':'11'}~~	~~fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8~~
2	6	{'name':'PHQ-9','score':'3'}	9d5aa6bb993c109599ef69669f2cd3484dfb8fa02e6879c6a9acf7936a537da0
· · ·		· · ·	
1	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d
3	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d

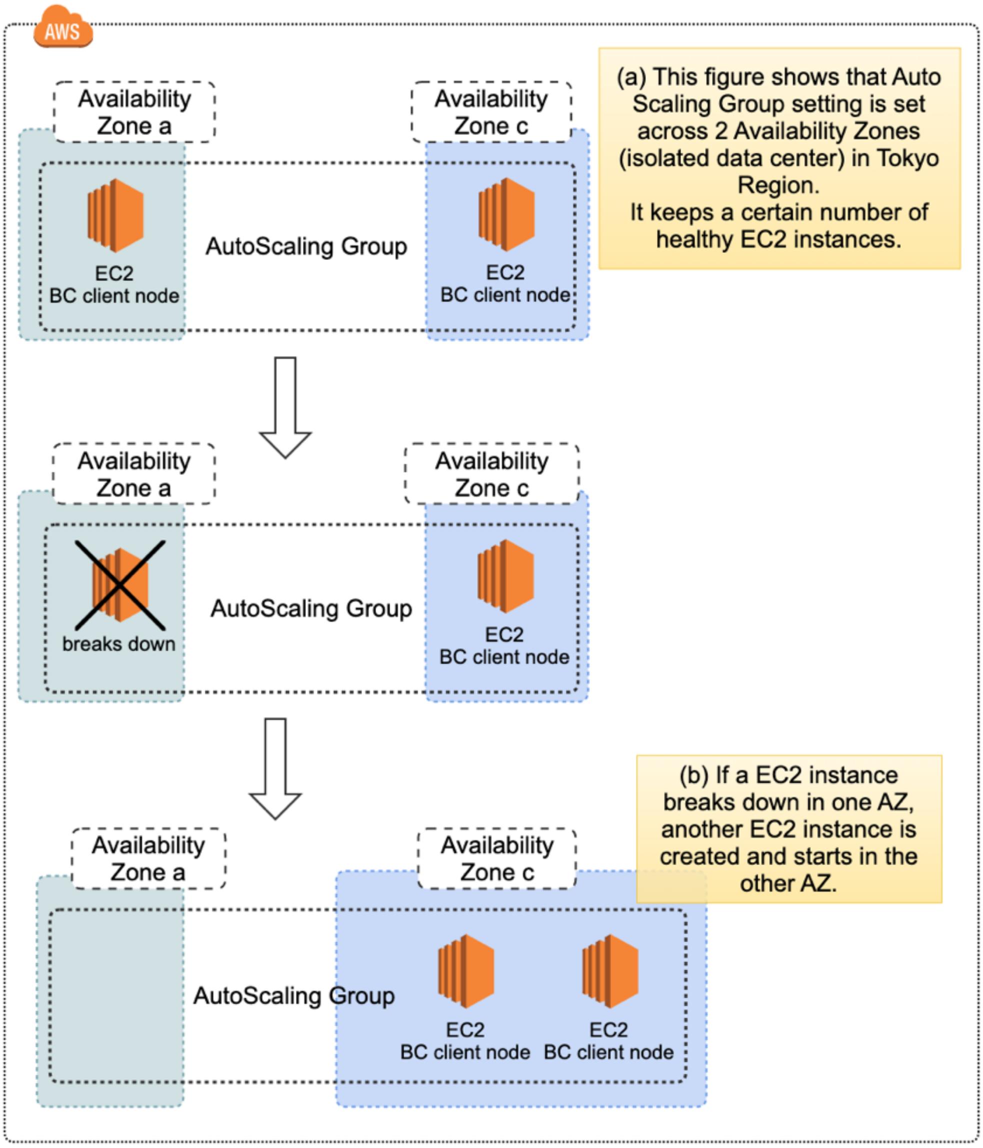
Check data chain

relay server	LogID	LogData	HashChain
· · ·		· · ·	
1	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
3	5	{'score':'11122','name':'EQ-5D'}	86aafed66f4d7a8f73028be00750877211b6a5460b85b681c20ce025ceb5cb22
1	6	{'name':'PHQ-9','score':'11'}	fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8
2	6	~~{'name':'PHQ-9','score':'11'}~~	~~fe9b66f5dc6ade250058f9dece62265ca19b7a7da656621d7116f8073de42fc8~~
2	6	{'name':'PHQ-9','score':'3'}	9d5aa6bb993c109599ef69669f2cd3484dfb8fa02e6879c6a9acf7936a537da0
· · ·		· · ·	
1	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d
3	16	{'key':'iq4YI0YNu3w1oTUAdqAJ'}	e2e82988b701a79aead637bf3b0e2cdb6a7e6939652ede535108ae911b690d1d

Data Integrity During Amazon Web Services Disruption Event

During the trial, we encountered a disruption event of the AWS cloud server in the Tokyo region on August 23, 2019 [29]. This resulted in impaired EC2 instances for some resources in the affected area of the AZ. One of the blockchain network nodes shut down due to disruption of the data center. Since the blockchain network was composed of three organizations that contain two validating peers and we had configured the blockchain network to span multiple AZs in AWS, the redundancy maintained stable operation of the system during the event. Furthermore, our configuration of a periodic health check function using Amazon EC2 Auto Scaling Service automatically restored the affected node in a healthy data center (Figure 9). We have confirmed there were no error records during the disruption event.

Figure 9. System resilience during Amazon Web Services disruption event in the Tokyo region. To secure the redundancy of the system, we have configured the blockchain network to span multiple availability zones (AZs) in Amazon Web Services. Each AZ is made up of one or more data centers equipped with independent power, cooling, and networking to ensure fault tolerance. We have also implemented a periodic health check function using the Amazon EC2 Auto Scaling Service. When one of the nodes in the blockchain network broke down due to the severe disruption of the data center on August 23, 2019, our system maintained stable operation and lost no data by the redundancy of the blockchain network. Furthermore, the health check system using the Amazon EC2 Auto Scaling Group enabled the automatic recovery of the blockchain node in a healthy AZ after the data center disruption event.



(a) This figure shows that Auto Scaling Group setting is set across 2 Availability Zones (isolated data center) in Tokyo Region.
It keeps a certain number of healthy EC2 instances.

(b) If a EC2 instance breaks down in one AZ, another EC2 instance is created and starts in the other AZ.

Even if further instances in the blockchain network or relay server components break down, we are able to prevent data loss. There are queues in the client smartphone app and relay servers, and the queues retain the clinical data until data are successfully sent. While the queues contain the data, we can recover the affected system.

Discussion

Principal Findings

In this study, we showed that a data management system using blockchain technology can reliably protect medical data and provide an immutable and fully traceable audit trail in a clinical trial. Due to the sensitivity of a hash function's output in relation to its input, any change in clinical data will result in a completely different hash string. This string will be an input of the next block's hash function, and the resulting string will be completely different from what it was prior to the data modification. As such, data integrity in a blockchain-based data management system can be checked by simply comparing the hash values of a blockchain. Verifying integrity can be done quickly, as the regulator needs only check for hash string equivalence. Our system can thus offer an improvement in clinical trial data management, enhancing trust in the clinical research process and easing regulator burden [23]. Although we used Hyperledger Fabric for the blockchain platform, other open-source platforms such as Ethereum can also be applied [42,43]. Furthermore, our system can also be used in clinical trials that collect data via other devices like PC or Internet of Things.

ICH-GCP provides guidance for monitoring conduct of a clinical trial to verify that reported clinical data are complete, accurate, and accounted for by source records. Representatives from the pharmaceutical and medical device industries must visit hospitals frequently and verify data by SDV to confirm that collected medical data are transferred correctly to the case report forms. It has been reported that almost half of the total budget for a phase 3 clinical trial is GCP-related and half of the GCP-related cost is related to SDV [1,3]. The total cost for current SDV in a typical phase 3 trial is estimated at US $90 million, which will be added to the price of the drug [3]. As the size and complexity of clinical trials grow, the cost of monitoring by the traditional method has become increasingly expensive. EMA and FDA have published updated guidelines for the monitoring and oversight of clinical trials. Both agencies encourage study sponsors to implement remote and risk-based monitoring [44,45]. Although remote monitoring may be an attractive alternative to traditional SDV, some studies have shown that remote monitoring can be more time-consuming than traditional monitoring and may be insufficient as a complete alternative for SDV [46]. Another option, risk-based monitoring, also needs risk assessment before its implementation and will not eliminate the need for data verification by the human eye. Conversely, our blockchain-based system enables secure data management without labor-intensive monitoring operations in medical practice. It can drastically reduce the cost of human resources and the risk of human-induced error. Information technology systems for the management of clinical trials need to keep audit trails permanent, tamper-proof, and verifiable, as stipulated by FDA 21 CFR part 11. Blockchain for clinical research involves a promising set of technologies that may advance data integrity and efficiencies in clinical research.

This study was conducted to demonstrate clinical data management using blockchain technology under certification of the regulatory sandbox of the Japanese Cabinet Office. The Japanese government aims to actively explore innovative and disruptive technologies through this regulatory sandbox system under the Cabinet Secretariat to address challenges such as extreme aging of the population [28]. Japanese society is facing challenges like increased national public health expenditures, higher demand for health care services lacking appropriate cost controls, and shortage of health care workers. Some of these challenges are becoming more acute, with recent data indicating that expenditures in health care insurance are growing in Japan [47]. The data collected through the demonstrations in the regulatory sandbox will be used in deliberation for regulatory reform, which will facilitate the creation of innovative business activities with new technologies and new business models.

Severe disruption of cloud server services occurred in the AWS Tokyo region on August 23, 2019, during the clinical trial. Thousands of companies, ranging from internet retailers and smartphone payment platforms to game providers, were affected and their services were suspended. During the event, customers of these services could not access their own information and lost data that should have been generated during the event. Our system maintained stable operation and lost no data from the participants of the clinical trial during the disruption event due to the redundancy of the blockchain network. The configuration of the blockchain network to span multiple data centers in AWS, along with the health check system using Amazon EC2 Auto Scaling, enabled automatic recovery of the blockchain network after severe disruption of the data center. These features in our system, which provide zero downtime for the availability of patient data, are suitable to meet the requirements of clinical trials.

Limitations

Further studies are needed to verify the scalability of the system for conducting various kinds of clinical trials. Although the transaction throughput in the Hyperledger Fabric platform that we used is much higher than public blockchain, further studies like multiple clinical trials in a single system may be fruitful. Data collection via multiple devices should also be tested toward the dissemination of virtual clinical trials.

Conclusion

In our study, we ran a clinical trial for breast cancer using a blockchain-based clinical data management system. The data were validated by protocol, and traceability was secured. Even with the severe disruption event of cloud server services, our system preserved the integrity of the clinical trial data. The study was conducted as a regulatory sandbox project by the Cabinet Office of Japan and will open the possibility for future deregulation measures enabling efficient clinical development.

Acknowledgments

This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO) and the Cabinet Office of Japan. We thank Kuniharu Tanabe and Satoshi Ota for the critical reading of the manuscript.

Authors' Contributions

TU designed the research; TH, TM, KO, KT, YM, and YO performed the research; TH and DI analyzed the data; and TH, TM, TK, DI, and TU wrote the paper.

Conflicts of Interest

TH, TM, KO, KT, TK, DI, and TU are members of SUSMED Inc. YM has received speaker fees from Suntory, Pfizer, Mochida, Eli Lilly, Morinaga Milk, and NTT Data and is conducting collaborative research with Morinaga Milk and SUSMED Inc.

References

1. Olsen R, Bihlet AR, Kalakou F, Andersen JR. The impact of clinical trial monitoring approaches on data integrity and cost: a review of current literature. Eur J Clin Pharmacol 2016 Apr;72(4):399-412. [doi: 10.1007/s00228-015-2004-y] [Medline: 26729259]
2. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 2012 Mar 1;11(3):191-200. [doi: 10.1038/nrd3681]
3. Funning S, Grahnén A, Eriksson K, Kettis-Linblad A. Quality assurance within the scope of Good Clinical Practice (GCP)—what is the cost of GCP-related activities? A survey within the Swedish Association of the Pharmaceutical Industry (LIF)'s members. Qual Assur J 2009 Jan;12(1):3-7 [FREE Full text] [doi: 10.1002/qaj.433]
4. Bhatt A. Quality of clinical trials: a moving target. Perspect Clin Res 2011 Oct;2(4):124-128 [FREE Full text] [doi: 10.4103/2229-3485.86880] [Medline: 22145122]
5. Gupta A. Fraud and misconduct in clinical research: a concern. Perspect Clin Res 2013 Apr;4(2):144-147 [FREE Full text] [doi: 10.4103/2229-3485.111800] [Medline: 23833741]
6. George SL, Buyse M. Data fraud in clinical trials. Clin Investig (Lond) 2015;5(2):161-173 [FREE Full text] [doi: 10.4155/cli.14.116] [Medline: 25729561]
7. McCurry J. Former Novartis employee arrested over valsartan data. Lancet 2014 Jun 21;383(9935):2111 [FREE Full text] [Medline: 24959638]
8. Yanagawa H, Nokihara H, Yokoi H, Houchi H, Nagai M, Yamashita R, Shikoku Collaborative Group for Promotion of Clinical Trials. Present status and perspectives on future roles of Japanese clinical research coordinators. J Clin Med Res 2018 Dec;10(12):877-882 [FREE Full text] [doi: 10.14740/jocmr3602w] [Medline: 30425759]
9. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. 2008. URL: https://bitcoin.org/bitcoin.pdf [accessed 2020-05-06]
10. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220. [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
11. Raghavendra M. Can blockchain technologies help tackle the opioid epidemic: a narrative review. Pain Med 2019 Oct 01;20(10):1884-1889. [doi: 10.1093/pm/pny315] [Medline: 30848821]
12. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention. JMIR Res Protoc 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: 10.2196/10163] [Medline: 30213780]
13. Tseng J, Liao Y, Chong B, Liao S. Governance on the drug supply chain via gcoin blockchain. Int J Environ Res Public Health 2018 Dec 23;15(6):1 [FREE Full text] [doi: 10.3390/ijerph15061055] [Medline: 29882861]
14. Zhou L, Wang L, Sun Y. MIStore: a blockchain-based medical insurance storage system. J Med Syst 2018 Jul 02;42(8):149 [FREE Full text] [doi: 10.1007/s10916-018-0996-4] [Medline: 29968202]
15. Wang H, Song Y. Secure cloud-based EHR system using attribute-based cryptosystem and blockchain. J Med Syst 2018 Jul 05;42(8):152. [doi: 10.1007/s10916-018-0994-6] [Medline: 29974270]
16. Mayer AH, da Costa CA, Righi RDR. Electronic health records in a blockchain: a systematic review. Health Informatics J 2019 Sep 30:1460458219866350. [doi: 10.1177/1460458219866350] [Medline: 31566472]
17. Beinke JH, Fitte C, Teuteberg F. Towards a stakeholder-oriented blockchain-based architecture for electronic health records: design science research study. J Med Internet Res 2019 Oct 07;21(10):e13585 [FREE Full text] [doi: 10.2196/13585] [Medline: 31593548]
18. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Blockchain vehicles for efficient medical record management. NPJ Digit Med 2020 Jan 6;3(1):3. [doi: 10.1038/s41746-019-0211-0]
19. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]
20. Angeletti F, Chatzigiannakis I, Vitaletti A. The role of blockchain and IoT in recruiting participants for digital clinical trials. 2017 Presented at: 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM); 2017; Split. [doi: 10.23919/softcom.2017.8115590]
21. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: Transparency and traceability of consent. F1000Res 2017;6:66 [FREE Full text] [doi: 10.12688/f1000research.10531.5] [Medline: 29167732]

22. Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T. Secure and scalable mhealth data management using blockchain combined with client hashchain: system design and validation. J Med Internet Res 2019 May 16;21(5):e13385 [FREE Full text] [doi: 10.2196/13385] [Medline: 31099337]
23. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]
24. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
25. Osipenko L. Blockchain’s potential to improve clinical trials: an essay by Leeza Osipenko. BMJ 2019 Oct 03:l5561. [doi: 10.1136/bmj.l5561]
26. Wust K, Gervais A. Do you need a blockchain? 2018 Presented at: 2018 Crypto Valley Conference on Blockchain Technology (CVCBT); 2018; Zug. [doi: 10.1109/cvcbt.2018.00011]
27. Mackey T, Bekki H, Matsuzaki T, Mizushima H. Examining the potential of blockchain technology to meet the needs of 21st-century Japanese health care: viewpoint on use cases and policy. J Med Internet Res 2020 Jan 09;22(1):e13649 [FREE Full text] [doi: 10.2196/13649] [Medline: 31917371]
28. Japan External Trade Organization. New regulatory sandbox framework in Japan. 2018 Jun 06. URL: https://www.jetro.go.jp/ext_images/en/invest/incentive_programs/pdf/Detailed_overview.pdf [accessed 2020-02-23]
29. Summary of the Amazon EC2 and Amazon EBS service event in the Tokyo (ap-northeast-1) region. 2019. URL: https://aws.amazon.com/jp/message/56489/ [accessed 2020-02-23]
30. Tsuji K, Ochi E, Okubo R, Shimizu Y, Kuchiba A, Ueno T, et al. Effect of home-based high-intensity interval training and behavioural modification using information and communication technology on cardiorespiratory fitness and exercise habits among sedentary breast cancer survivors: habit-B study protocol for a randomised controlled trial. BMJ Open 2019 Aug 22;9(8):e030911 [FREE Full text] [doi: 10.1136/bmjopen-2019-030911] [Medline: 31444192]
31. Androulaki E, Manevich Y, Muralidharan S, Murthy C, Nguyen B, Sethi M. Hyperledger fabric. 2018 Presented at: Proc Thirteenth EuroSys Conf EuroSys; 2018; Porto Portugal p. 1-15. [doi: 10.1145/3190508.3190538]
32. Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the global physical activity questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. BMC Public Health 2014 Dec 10;14:1255 [FREE Full text] [doi: 10.1186/1471-2458-14-1255] [Medline: 25492375]
33. Chu AHY, Ng SHX, Koh D, Müller-Riemenschneider F. Reliability and validity of the self- and interviewer-administered versions of the global physical activity questionnaire (GPAQ). PLoS One 2015;10(9):e0136944 [FREE Full text] [doi: 10.1371/journal.pone.0136944] [Medline: 26327457]
34. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. J Phys Act Health 2009 Nov;6(6):790-804. [Medline: 20101923]
35. Vickberg SMJ. The Concerns About Recurrence Scale (CARS): a systematic measure of women's fears about the possibility of breast cancer recurrence. Ann Behav Med 2003;25(1):16-24. [Medline: 12581932]
36. Momino K, Akechi T, Yamashita T, Fujita T, Hayahi H, Tsunoda N, et al. Psychometric properties of the Japanese version of the Concerns About Recurrence Scale (CARS-J). Jpn J Clin Oncol 2014 May;44(5):456-462. [doi: 10.1093/jjco/hyu032] [Medline: 24711631]
37. Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Yoshida M, Otsubo T, et al. The patient health questionnaire, Japanese version: validity according to the mini-international neuropsychiatric interview-plus. Psychol Rep 2007 Dec;101(3 Pt 1):952-960. [doi: 10.2466/pr0.101.3.952-960] [Medline: 18232454]
38. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res 2000 Jun;48(6):555-560. [doi: 10.1016/s0022-3999(00)00095-7] [Medline: 11033374]
39. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [Medline: 11491192]
40. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002 Jun;11(4):341-353. [doi: 10.1002/hec.673] [Medline: 12007165]
41. Gueron S, Johnson S, Walker J. SHA-512/256. 2011 Proc Eighth Int Conf Inf Technol New Gen 2011:354-358. [doi: 10.1109/itng.2011.69]
42. Chowdhury MJM, Ferdous MS, Biswas K, Chowdhury N, Kayes ASM, Alazab M, et al. A comparative analysis of distributed ledger technology platforms. IEEE Access 2019;7:167930-167943 [FREE Full text] [doi: 10.1109/access.2019.2953729]
43. Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 01;26(5):462-478. [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
44. Food and Drug Administration. Oversight of clinical investigations? A risk-based approach to monitoring. 2013. URL: https://www.fda.gov/media/116754/download [accessed 2020-02-23]
45. European Medicines Agencies. Reflection paper on risk based quality management in clinical trials. 2013. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials_en.pdf [accessed 2020-02-23]

46. Mealer M, Kittelson J, Thompson BT, Wheeler AP, Magee JC, Sokol RJ, et al. Remote source document verification in two national clinical trials networks: a pilot study. PLoS One 2013;8(12):e81890 [FREE Full text] [doi: 10.1371/journal.pone.0081890] [Medline: 24349149]
47. Organisation for Economic Co-operation and Development. 2019. Health at a Glance 2019 OECD Indicators URL: https://www.oecd-ilibrary.org/deliver/4dd50c09-en.pdf?itemId=%2Fcontent%2Fpublication%2F4dd50c09-en&mimeType=pdf [accessed 2020-05-04]

Abbreviations

AWS: Amazon Web Services
AZ: availability zone
EMA: European Medicines Agency
FDA: US Food and Drug Administration
ICH-GCP: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use–Good Clinical Practice
PMDA: Pharmaceuticals and Medical Devices Agency
SDV: source data verification
SQS: Simple Queue Service

Edited by G Eysenbach, K Clauson, P Zhang; submitted 28.03.20; peer-reviewed by L Osipenko, YR Park, JT te Gussinklo, E Chukwu, C Reis, TT Kuo, C Esposito; comments to author 20.04.20; revised version received 27.04.20; accepted 27.04.20; published 02.06.20.

Please cite as:
Hirano T, Motohashi T, Okumura K, Takajo K, Kuroki T, Ichikawa D, Matsuoka Y, Ochi E, Ueno T
Data Validation and Verification Using Blockchain in a Clinical Trial for Breast Cancer: Regulatory Sandbox
J Med Internet Res 2020;22(6):e18938
URL: https://www.jmir.org/2020/6/e18938
doi:10.2196/18938
PMID:32340974

Corrigenda and Addenda

Correction: Considerations for an Individual-Level Population Notification System for Pandemic Response: A Review and Prototype

Mohammad Nazmus Sakib[1], MBBS, MSc; Zahid A Butt[1], MBBS, MSc, PhD; Plinio Pelegrini Morita[1,2,3,4,5], MSc, PhD, PEng; Mark Oremus[1], PhD; Geoffrey T Fong[1,6,7], PhD, FRSC, FCAHS; Peter A Hall[1], PhD

[1]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada
[2]Research Institute for Aging, University of Waterloo and Schlegel Villages, Waterloo, ON, Canada
[3]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada
[4]eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
[5]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[6]Department of Psychology, University of Waterloo, Waterloo, ON, Canada
[7]Ontario Institute for Cancer Research, Toronto, ON, Canada

Corresponding Author:
Peter A Hall, PhD
School of Public Health and Health Systems
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 38110
Email: pahall@uwaterloo.ca

Related Article:

Correction of: https://www.jmir.org/2020/6/e19930/

(J Med Internet Res 2020;22(6):e21634) doi:10.2196/21634

In “Considerations for an Individual-Level Population Notification System for Pandemic Response: A Review and Prototype” (J Med Internet Res 2020;22(6):e19930), a correction request received from the authors during proofreading was inadvertently missed by JMIR Publications staff before publication.

The first sentence of the Introduction section, the word “organisms” should have been corrected to “pathogens.” This text has now been corrected from:

> *“Despite a dramatic improvement in health care affordances and extensive public health measures around the world, the emergence and re-emergence of infectious organisms and associated diseases have become a common phenomenon.”*

to:

> *“Despite a dramatic improvement in health care affordances and extensive public health measures around the world, the emergence and re-emergence of infectious pathogens and associated diseases have become a common phenomenon.”*

The correction will appear in the online version of the paper on the JMIR Publications website on June 22, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories

Submitted 19.06.20; this is a non–peer-reviewed article;accepted 19.06.20; published 22.06.20.

Please cite as:
Sakib MN, Butt ZA, Morita PP, Oremus M, Fong GT, Hall PA
Correction: Considerations for an Individual-Level Population Notification System for Pandemic Response: A Review and Prototype
J Med Internet Res 2020;22(6):e21634
URL: http://www.jmir.org/2020/6/e21634/
doi:10.2196/21634
PMID:32568723

Corrigenda and Addenda

Correction: Offline Digital Education for Postregistration Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration

Pawel Posadzki[1], PhD; Malgorzata M Bala[2], MD, PhD; Bhone Myint Kyaw[3], PhD, MBBS; Monika Semwal[1], PhD; Ushashree Divakar[1], MSc; Magdalena Koperny[4], PhD; Agnieszka Sliwka[5], PhD; Josip Car[1], MD, PhD, FRCP, FFPH

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]Chair of Epidemiology and Preventive Medicine, Department of Hygiene and Dietetics, Jagiellonian University Medical College, Krakow, Poland
[3]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[4]Province Sanitary-Epidemiological Station of Lesser Poland, Public Health and Health Promotion Department, Krakow, Poland
[5]Institute of Physiotherapy, Faculty of Health Sciences, Jagiellonian University Medical College, Krakow, Poland

Corresponding Author:
Josip Car, MD, PhD, FRCP, FFPH
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road
Singapore
Singapore
Phone: 65 6904 7005
Email: josip.car@ntu.edu.sg

Related Article:

Correction of: https://www.jmir.org/2019/4/e12968/

(J Med Internet Res 2020;22(6):e20316) doi:10.2196/20316

The authors of “Offline Digital Education for Postregistration Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration” (J Med Internet Res 2019;21(4):e12968) noticed an error in one of their author affiliations.

The affiliation for Malgorzata M Bala has been changed from:

> *2Health Services Department, The Agency for Health Technology Assessment and Tariff System, Warsaw, Poland*

to:

> *2Chair of Epidemiology and Preventive Medicine, Department of Hygiene and Dietetics, Jagiellonian University Medical College, Krakow, Poland*

The correction will appear in the online version of the paper on the JMIR website on June 23, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 15.05.20; this is a non–peer-reviewed article;accepted 15.05.20; published 23.06.20.

Please cite as:
Posadzki P, Bala MM, Kyaw BM, Semwal M, Divakar U, Koperny M, Sliwka A, Car J
Correction: Offline Digital Education for Postregistration Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration
J Med Internet Res 2020;22(6):e20316
URL: https://www.jmir.org/2020/6/e20316
doi:10.2196/20316
PMID:32574149

Viewpoint

Digital Health Equity and COVID-19: The Innovation Curve Cannot Reinforce the Social Gradient of Health

Allison Crawford[1,2], MD, PhD, FRCP; Eva Serhal[1], MBA, PhD

[1]Virtual Mental Health and Outreach, Centre for Addiction and Mental Health, Toronto, ON, Canada

[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Allison Crawford, MD, PhD, FRCP
Virtual Mental Health and Outreach
Centre for Addiction and Mental Health
250 College St
Room 825
Toronto, ON, M5T1R8
Canada
Phone: 1 416 659 5591
Email: allison.crawford@utoronto.ca

Abstract

Digital health innovations have been rapidly implemented and scaled to provide solutions to health delivery challenges posed by the coronavirus disease (COVID-19) pandemic. This has provided people with ongoing access to vital health services while minimizing their potential exposure to infection and allowing them to maintain social distancing. However, these solutions may have unintended consequences for health equity. Poverty, lack of access to digital health, poor engagement with digital health for some communities, and barriers to digital health literacy are some factors that can contribute to poor health outcomes. We present the Digital Health Equity Framework, which can be used to consider health equity factors. Along with person-centered care, digital health equity should be incorporated into health provider training and should be championed at the individual, institutional, and social levels. Important future directions will be to develop measurement-based approaches to digital health equity and to use these findings to further validate and refine this model.

(J Med Internet Res 2020;22(6):e19361) doi:10.2196/19361

KEYWORDS

virtual health; digital determinants of health; digital health equity; digital health; equity; COVID-19; public health; eHealth; social

Introduction

The public health crisis posed by coronavirus disease (COVID-19) has ignited rapid implementation of digital health care. In this commentary, we urge the implementation of health equity–informed digital health solutions. We introduce the Digital Health Equity Framework (DHEF) to identify the digital determinants of health and their links to digital health equity. In the current response to the COVID-19 pandemic, digital health has been rightly heralded as an innovative health solution that can ensure ongoing access to clinical care and allow public health measures that stem rapid viral transmission and spread [1,2]. However, unexamined inequities in access to and implementation of digital health as well as the quality of care afforded by digital health can recapitulate and deepen the inequities that have long existed within our health care system.

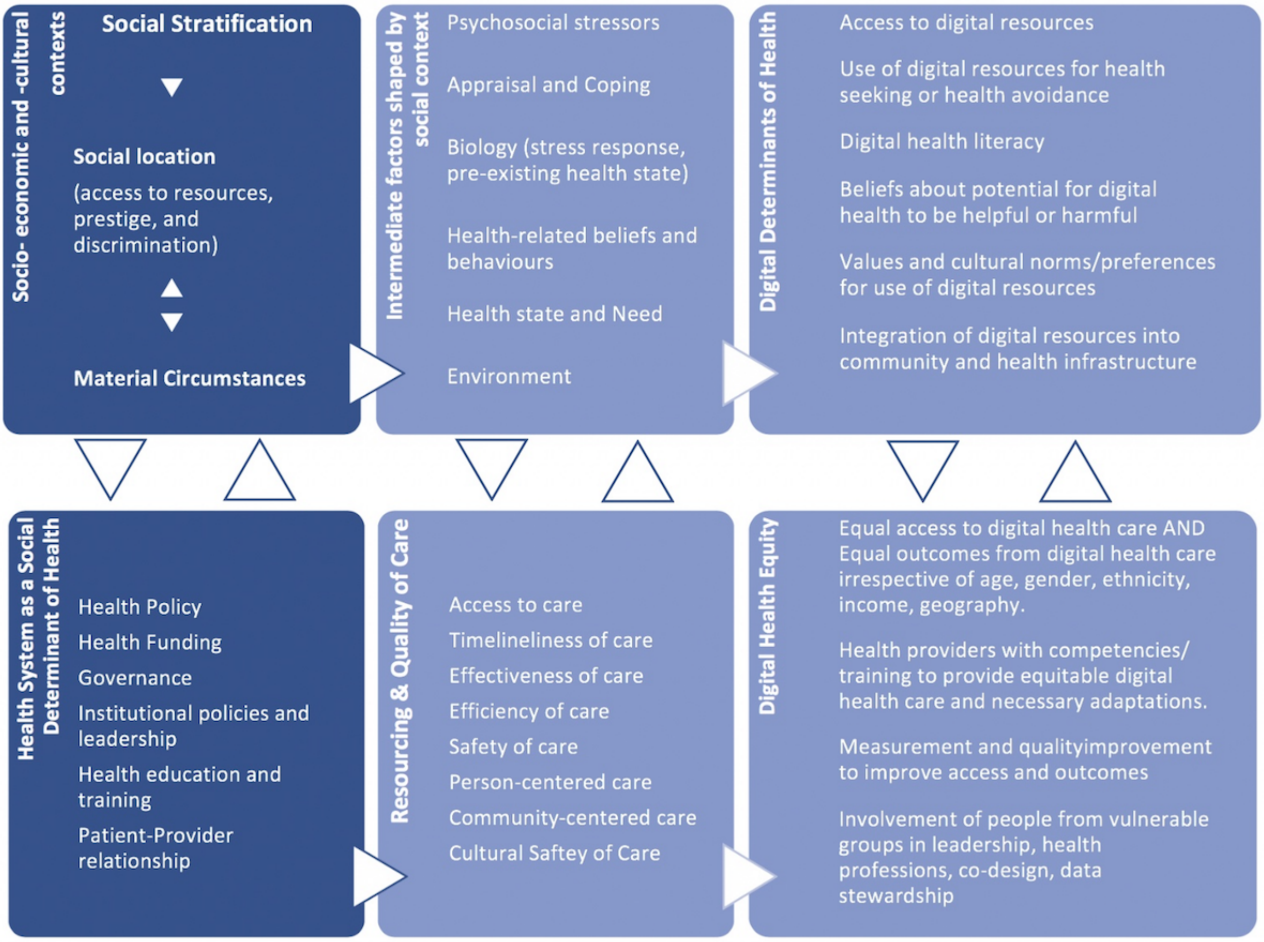
Digital health is broadly defined as "the field of knowledge and practice associated with the development and use of digital technologies to improve health" [3] across the full range of health technologies introduced into care, including telehealth, mobile health apps and wearable technologies, and online health services and tools. During the COVID-19 public health crisis, two modes of digital health have been commonly used: virtual health care, or televideo-enabled interactions between health providers and patients, and health information that is accessed online or via mobile apps.

Some media commentators [4] have stated that COVID-19 is the "great leveler" because it knows no boundaries and can infect rich and poor, young and old. However, this uncritical perspective misses the systemic factors that impact outcomes of illness and create health inequities between communities and across the life courses of individuals. Mounting evidence suggests that the COVID-19 pandemic has far greater associated morbidity and mortality in racialized groups that struggle with poverty and poor access to health care; the pandemic has also been suggested to compound preexisting inequities [5]. Similarly, there has been a lack of attention to health equity in the development of digital health solutions [6]; therefore, when these solutions are applied within the pandemic response, they may have unintended consequences of furthering health inequity. For example, access to technology can be limited by poverty, under-resourcing of health systems and neighborhoods, homelessness, and other factors that decrease engagement with technology and with digital health literacy skills. Health providers may also lack training and competencies in consideration of digital health equity as well as the cultural humility to understand how their patients and communities may experience or interact with technology. Digital health technologies interact with social, cultural, and economic realities and with social determinants of health to indirectly contribute to health equity.

The Digital Health Equity Framework (DHEF)

Here, we propose the DHEF (Figure 1), which applies many health equity factors outlined by Dover and Belon [7] in 2019, integrated with digital determinants of health and digital health equity. In their recent approach, Dover and Belon survey the many frameworks proposed for considering social determinants of health and address some of the limitations of the health equity field by moving from a superficial description of factors to a more comprehensive, ecological approach that considers the multitude of social, cultural, and economic factors that impact health and well-being as well as the interactions among these factors. Most significantly, they link these social determinants to health equity and begin to delineate a structure through which health equity can be measured within organizations and at a health system level.

Figure 1. Digital Health Equity Framework.



In Dover and Belon's model, which informs the foundation of the DHEF, the process of social stratification within economic and cultural social contexts refers to the hierarchical allocation and unequal distribution of power, prestige, and resources; this stratification assigns individuals to a social location, which is defined by intersectional factors such as race, age, income, geography, rurality, gender, ability, and occupation as well as other social factors. A person's social location governs exposure to health-related risks and vulnerabilities, including discrimination. A person's social location and material circumstances can be mutually reinforcing, and they also intersect with intermediate factors that shape health and health behaviors, including psychosocial stressors; styles of appraisal and coping; biology, including current and health status and preexisting conditions; health-related beliefs and behaviors; current health needs; and their environment [7].

Similarly, in the DHEF (Figure 1), *digital determinants of health* interact with other *intermediate health factors*, such as psychosocial stressors, preexisting health conditions, health-related beliefs and behaviors, and the environment, along with the person's current health state and needs. For example, access to digital health resources and digital health literacy interact with the degree and kind of psychosocial stress a person is currently experiencing; job loss or poverty, level of education, and previous exposure to digital media can all impact access. Styles of coping and appraisal of risk, along with health-related beliefs, can shape beliefs and behaviors regarding digital health; for example, some patients may have a tendency to avoid health care or to minimize risk, leading to issues such as corollary avoidance of digital health care, privacy-related concerns, or failure to appraise the quality of digital health information. Just as a person's environment shapes their health care access and quality, it also shapes their digital health access and quality; people living in overcrowded homes may lack privacy or, as with underhoused and homeless populations, may not have access to digital health solutions at all. All these intermediate factors are set into play, are reinforced, and in turn reinforce the *socioeconomic and sociocultural context* and the *social stratification* process. Intersections of race, gender, and geography are among the variables that determine one's power in society and define one's social location, which is closely linked to and interacts with one's material circumstances.

The DHEF expands on the *health system as a social determinant of health*. Moving the dial on health equity, including digital health equity, requires looking beyond individual factors to the health system. We need to ensure that at every level, from health care providers to institutions, insurers, health regulators, and government, we are able to detect, understand, and work to improve the *resourcing and quality of digital health care* for all social groups to reduce digital health disparities. Quality of care, which ensures that care is person-centered, safe, timely, effective, and efficient, is also care that is equitable [8]. This includes the quality of digital health care. For example, if digital health care is not experienced as culturally safe by a population of users, or if the environments (living spaces, communities, institutions, and infrastructure) and material circumstances of groups of people are not considered when developing institutional digital health strategies or in the provisioning of funding and remuneration models for providers working with vulnerable populations, the quality of digital care will suffer, and digital health equity will be impacted. The DHEF model highlights the importance of approaching digital health technologies from an ecological perspective, considering the ways that the use of technology by an individual extend out into (and are shaped by) their social, cultural, and economic position in the world. The case shown in Textbox 1 illustrates this interplay using COVID-19 as an example.

Textbox 1. Example of digital health equity related to COVID-19.

> Mr Seow is a 48-year-old man living with his older adult father and three children in an apartment in a socioeconomically disadvantaged urban neighborhood. His economic insecurity forces him to continue in his job as a food delivery worker; this exposes both Mr Seow and his father to higher risk of COVID-19 infection. Overcrowding in their apartment adds to that risk. Mr Seow has poorly controlled diabetes, a preexisting health condition that creates a risk of worse health outcomes with COVID-19 infection and is a consequence of the same social inequities. He also suffers from depression.
>
> Mr Seow's health risks related to COVID-19 may be further compounded by his own style of appraising risk and his own health behaviours. He tends to minimize risk, he smokes, and he has limited awareness of public health advice. His community has aging infrastructure, and the health resources in his neighborhood include an overcrowded hospital with inadequate access to lifesaving equipment such as ventilators.
>
> The effects of these social, cultural, and economic factors compound across Mr Seow's life course and include intergenerational cumulative effects of social location and material circumstances. Mr Seow's father, an older adult, is an immigrant who spent time in a detention center many years ago; he equates hospital with imprisonment, also avoids health care, and has many unmet health needs.
>
> The promise of digital health in relation to Mr Seow's situation is evident. Virtual visits (ie, telehealth) will allow him to continue to access health care for his diabetes and depression while enabling him to avoid exposure to the overcrowded hospital in his community. This has potential to mitigate some of his health risks and support health equity. However, digital determinants of health and likely outcomes of digital health equity should also be examined. Mr Seow's outcome of receiving virtual health may vary greatly compared with that of a man of the same age living in more economically advantaged circumstances. These inequities must be recognized to be addressed. For example, does Mr Seow have access to technology that supports virtual care? Can he meet with a health provider in a private space within his crowded living situation? Does he have a minimum degree of digital health literacy, with the ability to access and appraise reliable information? Do his own personal beliefs and values about technology support his use of digital health? Are digital health resources available in his community, and do they integrate with other points of access to health care in his community (ie, is this option covered by his insurer, or is it integrated with his primary care provider)?
>
> At the provider level, are Mr Seow's health care providers, such as doctors and nurses, trained to think about health equity? Do they possess the cultural humility to recognize some of these potential gaps and disadvantages of virtual care for Mr Seow and to make necessary adaptations? Are they trained to monitor outcomes of virtual health in their setting and to consider sociocultural variables in these outcomes?
>
> Addressing health equity extends far beyond the individual and the patient-provider dyad to systemic and social contexts. To ensure digital health equity, primary care, hospitals, and governments must have digital health strategies that identify and addresses potential gaps in digital health care based on these digital determinants of health. To know whether they are successful in achieving digital health equity, they need to conduct ongoing measurements of equity and of digital health outcomes. Health equity considerations must also be part of quality of care considerations, along with person-centered care that considers patient choice and autonomy.

There are examples, albeit few, of digital health research that incorporates considerations of social determinants and health equity, particularly in developing contexts [6]; however, this approach needs to become mainstream in all implementations of digital health. Only one week ago, the World Health Organization (WHO) released its 4-year draft global strategy on digital health [3], which aims to support international efforts "to develop the infrastructure for information and communication technologies for health…[and] to promote equitable, affordable and universal access to their benefits" along with promoting the development of national digital health strategies. In line with the 2030 Agenda for Sustainable Development, the WHO seizes on the "great potential to accelerate human progress, to bridge the digital divide and to develop knowledge societies." Within the WHO strategy, one of the strategic objectives is to advocate for people-centered health systems that are enabled by digital health. For example, they seek to advance "digital health literacy, gender equality and women's empowerment and inclusive approaches to adoption and management of digital health technologies." The report mentions a number of approaches that relate to health equity throughout; however, these are as yet unformulated within the WHO's implementation plan.

To consider digital health equity within our health and social contexts, we need to establish systematic ways to ensure that potential health inequities are identified and addressed in digital health policies, strategies, and programs so that existing health inequities are not reinscribed onto our virtual health landscapes. Implementation science, which specifies factors relevant to increased uptake of an innovation, is emerging as a critical factor to guide the spread and scaling of digital health; however, implementation models often fail to incorporate health equity factors or to address social determinants of health. To ensure health equity within digital health, we need to be purposeful in implementing digital health in an equitable way and in measuring health outcomes through an equity lens. If we do not collect health equity data, we cannot monitor health equity outcomes. In turn, understanding the population health needs of vulnerable groups can identify barriers to implementation that create innovation gaps, such as the gap between population health interest in digital health care and the capacity of US hospitals to deliver digital health care in response to COVID-19 [9]. Integrating a health equity approach such as the DHEF with a health implementation approach is an urgent need, particularly at this time of rapid advances in digital health innovation. This should include determining appropriate health equity metrics and measures for digital health.

Perhaps most importantly, to avoid duplicating the social stratification that exists in society at large, we need to ensure the meaningful involvement of people from marginalized and vulnerable groups in positions of digital health leadership, as health providers, and in codesign at all stages of innovation and implementation, including as stewards of their own health outcome data. In times of crisis, such as the current COVID-19 pandemic, utilitarian principles of innovation are often viewed as a way to maximize overall social benefits, while egalitarian principles that address inequalities are set aside [10].

Conclusion

There is no question that virtual health care can provide sustained access to essential health care; however, this commentary aims to draw attention to the unintended health equity impacts of the pivot to digital health care from the early days of the response to the current pandemic. One of the major limitations of this commentary is that we do not have available data to quantify these concerns. We believe that measurement-based approaches to health equity are a high priority for digital health research. There is emerging work in this area [11], and we hope that the framework we propose can further stimulate investigation into the multiple ways in which digital determinants of health may impact digital health equity. In turn, these data will lead to refinement of the proposed framework. During the COVID-19 pandemic and in its aftermath, we need to urge ongoing attention to health equity, including digital health equity, and to develop processes and measures to prevent our own blind spots and inattention in this regard. The celebrated curve of innovation cannot reinforce the social gradient of health, whereby people in less advantaged socioeconomic positions have less access to digital health care, poorer quality of digital health care, or worse health outcomes.

Conflicts of Interest

None declared.

References

1. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Apr 02:211-212. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
2. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
3. World Health Organization. 2020 Mar 22. Draft global strategy on digital health, 2020-2024 URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf) [accessed 2020-05-26]
4. Dooc E. Business Mirror. 2020 Mar 17. Opinion: Covid-19 is a great leveler URL: https://businessmirror.com.ph/2020/03/17/covid-19-is-a-great-leveler/ [accessed 2020-05-26]
5. Williams DR, Cooper LA. COVID-19 and Health Equity-A New Kind of "Herd Immunity". JAMA 2020 May 11:online ahead of print. [doi: 10.1001/jama.2020.8051] [Medline: 32391852]
6. Sinha C, Schryer-Roy A. Digital health, gender and health equity: invisible imperatives. J Public Health (Oxf) 2018 Dec 01;40(suppl_2):ii1-ii5 [FREE Full text] [doi: 10.1093/pubmed/fdy171] [Medline: 30329082]
7. Dover DC, Belon AP. The health equity measurement framework: a comprehensive model to measure social inequities in health. Int J Equity Health 2019 Feb 19;18(1):36 [FREE Full text] [doi: 10.1186/s12939-019-0935-0] [Medline: 30782161]

8. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press; 2001.
9. Hong Y, Lawrence J, Williams D, Mainous I. Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and National Hospital Survey Data. JMIR Public Health Surveill 2020 Apr 07;6(2):e18961 [FREE Full text] [doi: 10.2196/18961] [Medline: 32250963]
10. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/NEJMsb2005114] [Medline: 32202722]
11. Were MC, Sinha C, Catalani C. A systematic approach to equity assessment for digital health interventions: case example of mobile personal health records. J Am Med Inform Assoc 2019 Aug 01;26(8-9):884-890. [doi: 10.1093/jamia/ocz071] [Medline: 31188438]

Abbreviations

COVID-19: coronavirus disease
DHEF: Digital Health Equity Framework
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.04.20; peer-reviewed by D Hilty, N Schaeffeler; comments to author 21.05.20; revised version received 23.05.20; accepted 25.05.20; published 02.06.20.

Please cite as:
Crawford A, Serhal E
Digital Health Equity and COVID-19: The Innovation Curve Cannot Reinforce the Social Gradient of Health
J Med Internet Res 2020;22(6):e19361
URL: https://www.jmir.org/2020/6/e19361
doi:10.2196/19361
PMID:32452816

Original Paper

Online Information Exchange and Anxiety Spread in the Early Stage of the Novel Coronavirus (COVID-19) Outbreak in South Korea: Structural Topic Model and Network Analysis

Wonkwang Jo[1], PhD; Jaeho Lee[2], PhD; Junli Park[3], MD; Yeol Kim[2,3], MPH, MD, PhD

[1]The Institute for Social Data Science, Pohang University of Science and Technology, Pohang, Republic of Korea
[2]National Cancer Control Institute, National Cancer Center, Goyang, Republic of Korea
[3]Department of Family Medicine, National Cancer Center, Goyang, Republic of Korea

Corresponding Author:
Yeol Kim, MPH, MD, PhD
National Cancer Control Institute
National Cancer Center
323, Ilsan-ro, Ilsandong-gu
Goyang, 10408
Republic of Korea
Phone: 82 31 920 1753
Fax: 82 31 920 2189
Email: drheat@ncc.re.kr

Abstract

Background: In case of a population-wide infectious disease outbreak, such as the novel coronavirus disease (COVID-19), people's online activities could significantly affect public concerns and health behaviors due to difficulty in accessing credible information from reliable sources, which in turn causes people to seek necessary information on the web. Therefore, measuring and analyzing online health communication and public sentiment is essential for establishing effective and efficient disease control policies, especially in the early stage of an outbreak.

Objective: This study aimed to investigate the trends of online health communication, analyze the focus of people's anxiety in the early stages of COVID-19, and evaluate the appropriateness of online information.

Methods: We collected 13,148 questions and 29,040 answers related to COVID-19 from Naver, the most popular Korean web portal (January 20, 2020, to March 2, 2020). Three main methods were used in this study: (1) the structural topic model was used to examine the topics in the online questions; (2) word network analysis was conducted to analyze the focus of people's anxiety and worry in the questions; and (3) two medical doctors assessed the appropriateness of the answers to the questions, which were primarily related to people's anxiety.

Results: A total of 50 topics and 6 cohesive topic communities were identified from the questions. Among them, topic community 4 (suspecting COVID-19 infection after developing a particular symptom) accounted for the largest portion of the questions. As the number of confirmed patients increased, the proportion of topics belonging to topic community 4 also increased. Additionally, the prolonged situation led to a slight increase in the proportion of topics related to job issues. People's anxieties and worries were closely related with physical symptoms and self-protection methods. Although relatively appropriate to suspect physical symptoms, a high proportion of answers related to self-protection methods were assessed as misinformation or advertisements.

Conclusions: Search activity for online information regarding the COVID-19 outbreak has been active. Many of the online questions were related to people's anxieties and worries. A considerable portion of corresponding answers had false information or were advertisements. The study results could contribute reference information to various countries that need to monitor public anxiety and provide appropriate information in the early stage of an infectious disease outbreak, including COVID-19. Our research also contributes to developing methods for measuring public opinion and sentiment in an epidemic situation based on natural language data on the internet.

(J Med Internet Res 2020;22(6):e19455) doi:10.2196/19455

KEYWORDS

coronavirus; anxiety; pandemic; online; health information exchange; topic modeling

Introduction

The recent appearance of the novel coronavirus disease (COVID-19) has been devastating worldwide. In South Korea, hundreds of new cases have been diagnosed daily since late February 2020. The cumulative number of confirmed cases at the time of writing (April 1, 2020) exceeded 9000. Internationally, over 800,000 cases have been confirmed in more than 200 countries, areas, and territories [1], despite the World Health Organization's request for global efforts to slow down the spread of the virus the previous month [2]. Most countries have strongly recommended basic preventive methods such as quarantine and isolation of suspected cases, a macrolevel campaign on improving personal hygiene (eg, more frequent hand washing), or using masks in public sites. Additionally, some countries including Korea are now implementing more severe measures such as a social distancing that require the general population to refrain from congregating in public places.

In the event of a population-wide infectious disease outbreak such as COVID-19 people's online activities could significantly affect public concerns and health behaviors. Many studies have indicated people's active use of online information in various crisis situations, including a public health crisis [3,4]. Information and emotional exchanges between people on the internet form public opinions and concerns, which in turn affects people's cognition and behavior. Although these opinions and information on the internet are sometimes useful, they are not always appropriate. There could be dissemination of misinformation, which may lead to inappropriate medical advice or unnecessary anxiety [5-7].

Analyzing data on the internet that records how people voluntarily exchange opinions and information about COVID-19, such as relevant posts from social media services, provides a valuable opportunity to understand and monitor the public concerns over COVID-19 and the dissemination of related information on the internet. Considering the need to manage rumors and monitor public opinions and behaviors in the context of a mass infectious disease, and the importance of internet opinions in the event of a population-wide outbreak, the analysis of the online data has great implications for the formation of efficient and effective health policies and appropriate provision of information. These spontaneously written language materials contain a wealth of information about various topics on COVID-19 that cannot be thoroughly predicted by health policy makers and public health researchers, and thus cannot be measured by a traditional predetermined questionnaire. Hence, analyzing web-based data can supplement traditional surveys and contribute to making health policies for the general population [7-9].

Web data analysis is particularly valuable in the early stage of an infectious disease outbreak. In the early stage of a new disease outbreak, health authorities may lack proper guidelines for the disease, and people may not find trustworthy information from other sources. Because of this situation, people might be more affected by uncertain information on the internet. Thus, monitoring web data in the early stage of an infectious disease outbreak is important to prevent inappropriate dissemination of misinformation or unnecessary anxiety that could occur in the early stage of an outbreak.

This study primarily evaluated the public concerns over COVID-19 in the outbreak's early stage using data from the online social questions and answers (Q&A) forum in Korea's largest search engine, Naver.com [10,11], and analyzed the characteristics of item responses. The Naver Q&A forum (the service named "Jisik-In," meaning "an intellectual" in Korean) resembles Quora.com, as it allows users to post questions and answers on any topic. We used 13,148 questions and threaded answers posted on Naver's Q&A forum in the early stage of the COVID-19 outbreak to analyze the characteristics of the public's online concerns and the appropriateness of information circulated there.

In summary, our research questions include the following:

1. What is the focus of the questions on COVID-19 in the Naver Q&A forum?
2. How do the subjects observed in numerous questions change by time and main events?
3. What are the main objects concerning anxiety and worry on COVID-19?
4. How appropriate or significant is the information provided in the answers to the questions communicating anxiety and worry?

Methods

Data Collection

This study used the question and answer data available on Naver's Q&A forum. The forum is open to the public and allows individuals to post both questions and answers, anonymously or otherwise.

There were two main reasons for selecting Naver's Q&A forum among numerous online services available for public opinion exchange. First, Naver is a service provider with a dominant market position in Korea. It receives about 30 million visitors daily and is estimated to be used by about 76% of Korean Internet users as the main search portal website based on a recent survey [10,12]. Moreover, Naver is the only search engine of massive users with a Q&A forum that allows users to freely access the information exchange. Although similar information exchanges could happen in other internet communities and social network services, in general, only their members could approach and see them, so their influence is limited. Therefore, Naver's Q&A forum data better illustrates the Korean public's concern generated through online postings. Second, Unlike data from other social media, language data from a Q&A forum includes a detailed context of the author's interests and feelings. The question and answer form allows the user to post detailed information because its aim is to help others understand the full situation. A Twitter post, for example, often simply reveals an author's feelings or anxieties; however, a question from Naver's Q&A forum explains the issue's background. Our data, thus, may allow more informative analysis on general public concerns over COVID-19 than other sources of web-based data.

COVID-19–related questions posted between January 20, 2020, and March 2, 2020, and their respective answers were collected from Naver's Q&A forum and used for analysis. January 20, 2020, was chosen as the starting point for data collection since it corresponds with the diagnosis of the first patient with COVID-19 in Korea, and questions before this date are scarce. Duration of the previously mentioned data was then confirmed by considering the frequency and characteristics of the data during preprocessing. The procedure for identifying the COVID-19–related posts comprised several steps.

First, one question and its attached answers were considered as one document. We collected all the documents (questions and answers) containing the word "코로나" (in English, "corona") from December 30, 2019, to March 2, 2020. This, however, also extracts questions and answers using the word "코로나" that refers to objects other than COVID-19.

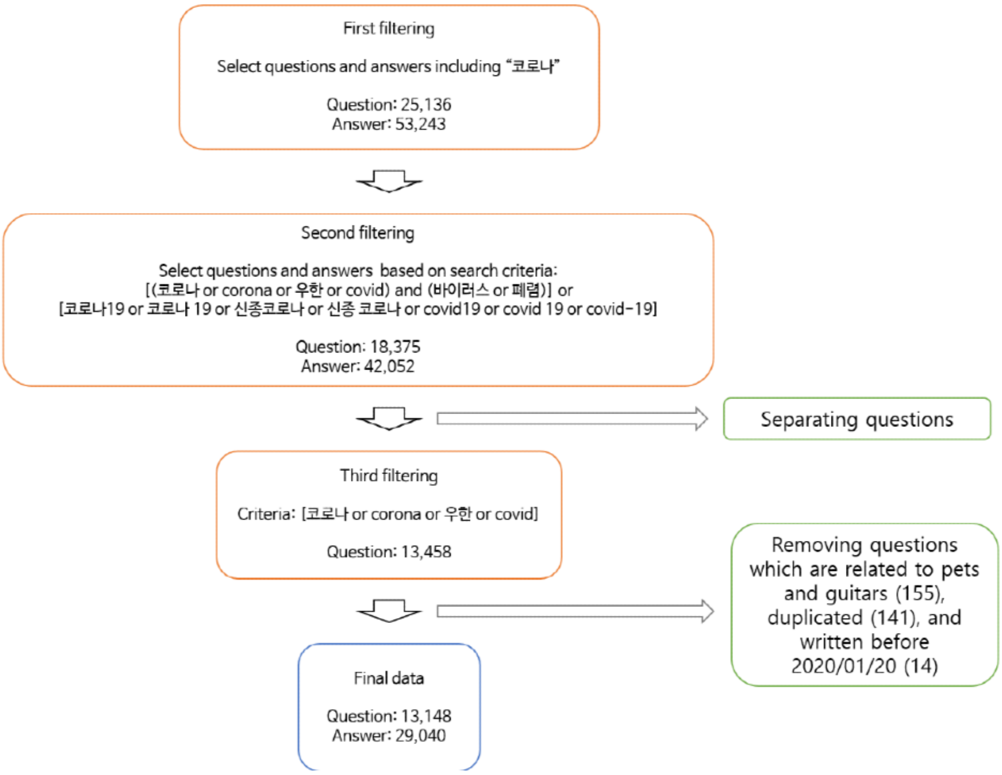
Second, the data were reselected from the results of the first step using additional search criteria to select COVID-19–related questions and answers more accurately. The search criteria were as follows: ([코로나 or corona or 우한 or COVID] and [바이러스 or 폐렴]) or (코로나19 or 코로나 19 or 신종코로나 or 신종코로나 or COVID19 or COVID 19 or COVID-19). English and Korean words were mixed in the search criteria because Koreans use both languages commonly. The English translations of the Korean words used here are as follows: 코로나=corona, 우한=Wuhan, 바이러스=virus, 폐렴=pneumonia, 신종코로나=novel corona. Therefore, we included biological words such as virus or pneumonia, or a relatively formal name referring to this disease, such as COVID19 or corona19, in the search criteria. These criteria were used to identify questions and answers that were related to COVID-19 and contained at least a fraction of biological perspective, as we assumed that the biological words and formal name reflect the perspective at least a little. Since COVID-19 is a controversial issue in domestic politics, we searched for posts that contained at least some biological perspective and excluded posts written purely from a political viewpoint.

Third, we isolated the questions from the selected data and again selected the questions satisfying the following search criteria: (코로나 or corona or 우한 or COVID). Questions including irrelevant words (eg, pets, guitars) were removed. Since dogs or cats could be infected with a different type of coronavirus than COVID-19, there were questions about it. We excluded the questions related to pets. Questions related to "corona," the guitar-producing company in Korea, were also excluded. The advanced criteria were applied only to the questions because some users supplied answers without considering the questions' contents. In other words, although rare, there were cases in which answers related to COVID-19 were attached to questions unrelated to COVID-19. This ensured that both questions and answers were related to COVID-19.

Fourth, the duplicate questions were deleted. Q&A forum users sometimes reposted the same question or posted a similar question with a slight change in the wording, which were unhelpful. The criteria used to find duplicate questions were whether the first 50 or the last 50 characters of the question, including blank spaces, were duplicated.

Fifth, the few questions (n=14) posted before January 20, 2020, were deleted, as the data in this period cannot represent public concern properly. As a result, 13,148 questions and 29,040 answers, which were assumed to be related to COVID-19, were collected. Figure 1 presents a schematic summary of this process.

Figure 1. Data filtering process. COVID-19: coronavirus disease.



Data Analysis

Several text-mining techniques, including structural topic modeling and word network analysis, were used to analyze public concerns from 13,148 questions. Language data analyses have often used human interpretation capability [13]. The theme in language materials results from synthesizing various information, more than what is simply and explicitly expressed. Therefore, it is convenient to capture themes by mobilizing human ability to interpret texts. However, there is a clear limitation on the amount of data that can be processed by a few researchers. This explains why several previous studies analyzing medical-related media or internet posts used a small amount of sample data [14-16]. These studies were also susceptible to the unfavorable effects of human researchers' subjectivity. Using text mining techniques to extract useful information from large volumes of data using computers, this study objectively estimated public concerns from large volumes of language data.

Although text mining allows us to examine a huge amount of data, these techniques usually cannot capture delicate nuances. For example, using basic text mining techniques, determining whether the answer is correct information or a rumor is somewhat difficult. This poses a challenge to the researcher because rumors, especially convincing rumors, use similar words and connection of words with valid information. To compensate for this, we also used a method that allows medical doctors (family medicine specialists) to categorize answers to questions on a particular topic and then analyze the characteristics of those answers.

In summary, three main methods were used for analysis: the structural topic model (STM), network analysis, and professional qualitative classification. Each method is described in the following sections.

Structural Topic Model

The STM, a type of topic modeling method, was selected to extract the overall themes or focus of 13,148 questions and examine how the theme or focus changed over time.

Like most topic modeling methods developed after the latent Dirichlet allocation (LDA), the STM can extract multiple topics and the topics' probability distribution in each document from a large number of documents. The extracted topics and their distribution are information that summarizes the given documents [17,18].

The topic estimation process is based on several assumptions. Topic modeling methods assume that a document is a simple set of words and a topic is a probability distribution of words (eg, cat: 0.015, dog: 0.01, pet: 0.009, etc). Each document contains multiple topics with specific probability distributions (eg, the first document: topic 1=0.4, topic 2=0.2, and topic 3=0.4). It is then assumed that individual documents were randomly generated from the topics and their distribution per document, and were not directly written by humans; thus, the most probable topics and their distributions are estimated considering the given data [17-19].

Naturally, the probability distribution of words itself does not have intuitive meaning; however, we can interpret the meaning of the topic from the probability distribution of words. When a topic is expressed qualitatively, it mainly appears as an unequal use of words. For example, suppose a topic of "cancer screening test" exists in the language materials. Certain words like "screen" or "mammography" would be used more frequently in this material than other words. Therefore, if we can deduce the word probability distributions, which are likely to produce the given documents, we can also deduce the topics' meanings by noting the high probability words in the corresponding probability distributions.

Distributions of topics in each document are also important information for interpreting the topics because they can be used to identify how each topic is realized into language material. We can also identify documents that have the highest proportion on each individual topic. For example, we can identify the top ten documents with the highest proportion of topic 2 among all documents. After reading the ten documents, we can understand the detailed context and intuitive meaning of topic 2 more accurately.

In short, the topic modeling method estimates topics and their probability distributions per document that explain the given documents appropriately. Although the probability distribution itself does not provide intuitive meaning, researchers could interpret the topics' meaning.

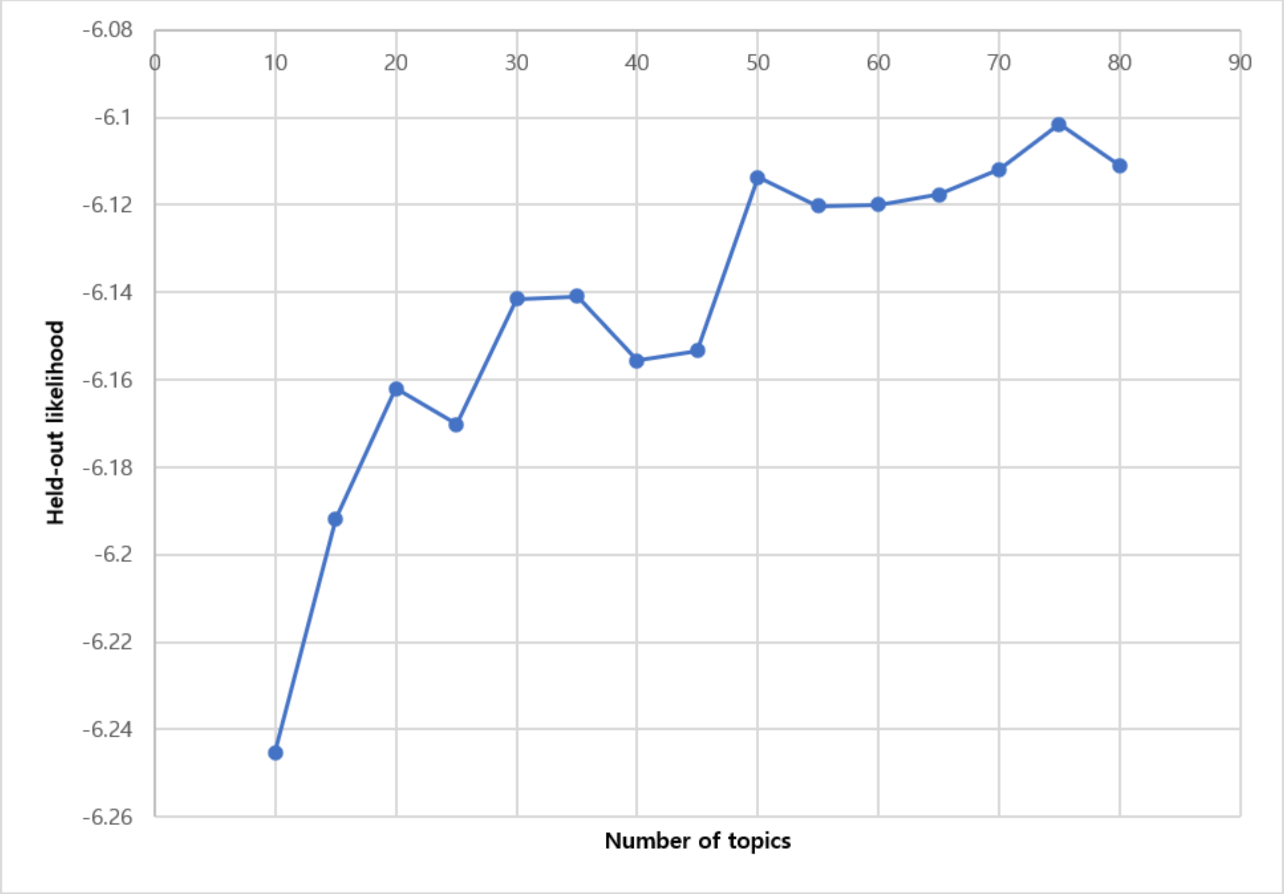
Besides this general function of topic modeling, the STM estimates how much of the document's meta information affects the topics' proportion or content [20,21]. Meta information refers to other information that exists in the document outside the document's content (eg, when a document was written or the type of author). The STM estimates how the meta information affects the proportion and content of the extracted topics. Given that this study primarily aimed to analyze how the topics of questions changed over time, these attributes of the STM were deemed appropriate to achieve our research objectives. This study estimated how the time of posting the question affected the proportion of question topics.

The application of the STM to the questions is described as follows. Of the 13,148 questions, 12 questions were additionally eliminated in the preprocessing for the STM. We only used the words that appeared in at least 2 questions and questions that contained at least 2 words, as a word that only appears in a single document or a question that contains only 1 word carries little information for topic modeling. We estimated 50 topics from the remaining 13,136 questions. The number of topics was set at 50 because there was no significant improvement in topic modeling performance after 50 as measured by the held-out likelihood [22] when changing the number of topics from 10 to 80 in increments of 5 (Figure 2). The posting time of the questions, which was measured by the unit of 1 day, was used as a covariate in our model for estimating topics' proportion change by time.

All the topics were interpreted and labeled by considering three kinds of information: the words given high probability in each topic, the words with a high frequency and exclusivity (FREX) score in each topic, and the documents with a high proportion of each topic. The importance of the high probability words and the documents estimated to have a high proportion of each topic in interpreting topics has been explained previously. The words with a high FREX score supplement the high probability words by considering exclusivity and frequency together [21]. That is, a high FREX score word of a topic is important, especially in the topic. All the authors collectively interpreted and labeled 50 topics, considering the top 15 probability words, the top 15 FREX score words, and the top 10 questions with a high proportion of each topic.

We also used the STM to sample representative questions containing subjects related to anxiety and worry. To analyze the appropriateness of the answers to the questions related to the public's anxiety and worry, most representative questions needed to be reselected from the entire batch of questions. Since it is impossible for a small group of human researchers to review 13,148 questions directly, the STM results were used to select questions that contained anxiety- and worry-related topics. We extracted questions that contained a high proportion of topics discussing physical symptoms and self-protection methods against COVID-19 because they were revealed as the main targets of anxiety and worry.

Figure 2. Held-out likelihood.



Network Analysis

Topic modeling is useful for identifying broad topics in a large volume of documents; however, a researcher cannot control the model to see results on specific subjects. Therefore, analyzing the relationship between particular words of interest is important in achieving our research goals.

One of our research objectives was to identify the objects and contexts of anxiety and worry expressed by people. To identify the source of anxiety and worry, we extracted the top 20 words that appeared most frequently in the questions including the words “불안” or “걱정,” which are Korean words for “anxiety” and “worry,” respectively. We considered both words appearing in the same question as a linkage or an association between the words.

Additionally, we created a word network using the top 50 words linked to the two Korean words referring to anxiety and worry to analyze the context of people’s anxiety and worry comprehensively. Observing individual words associated with anxiety and worry cannot accurately analyze the full context of anxiety and worry. A network of words linked to anxiety and worry allowed us to further analyze the context of anxiety and worry. Therefore, the top 50 words associated with the two Korean words referring to anxiety and worry were gathered, and the network was created. Connection criterion is a coappearance of words in the same question. This network is extremely dense because the network is made of words that are used in a similar context. Only the most prominent links were required to be selected to extract particularly prominent meanings. For this, only the links with the highest weight, (ie, 500 of the most frequent connections) were extracted to create a subnetwork. This subnetwork was assumed to contain the most prominent contexts related to people’s anxiety and worry.

The network community detection algorithm was applied to this subnetwork to extract distinguished themes from the network. This algorithm identifies relatively more cohesive communities of nodes within a network [23-25]. When applied to our word network, it worked to find a set of words that appeared more frequently with each other in the entire network. For example, the algorithm could judge that a “China,” “Wuhan,” “pneumonia,” “infection,” and “travel” word cluster appeared frequently in the entire word network. This was interpreted as a distinct theme that we found through network analysis. In other words, the network community detection algorithm identified the most prominent and distinct contexts that appeared when people expressed anxiety or worry. We identified cohesive communities of words, which contained at least 5 words because this amount was required to interpret the word community into a meaningful theme.

Walktrap was selected from the various network community detection algorithms because of its excellent performance while overcoming the “resolution problem” [25,26]. A resolution problem refers to a situation in which algorithms do not capture a community of a small number of nodes properly, a frequently faced problem in applying network community detection algorithms. Put simply, the Walktrap algorithm calculates the distance between nodes in the network using a random walk from each node and finds communities based on that distance. The analyst sets the step of the random walk for distance calculation, which we set to 2. The network analysis was implemented through the igraph package of R (R Foundation for Statistical Computing).

The Walktrap algorithm was also used to identify cohesive communities of topics related to each other. The STM estimates the correlation between topics [20,27]. A positive correlation between two topics means that the two topics are likely to appear together in the same document. We assumed this positive correlation to be a link between topics and formed a network of all topics. Next, using the network community detection algorithm, we analyzed whether there were relatively more cohesive communities of topics in the entire network of topics. In other words, we tried to identify sets of topics that were often expressed together. Consequently, we identified 6 topic communities that were judged to be cohesive. We labeled each topic community to express broader themes embracing topics belonging to the topic community, considering each topic's interpretation made through the STM.

That is, after estimating 50 topics from 13,136 questions, these topics were summarized into 6 topic communities using the Walktrap algorithm and then interpreted. This additional step was employed due to the fact that, although 50 topics is a great summary, it is still a lot of information for a person to grasp intuitively.

Professionals' Qualitative Coding

Using the results of the aforementioned methods, we assessed the appropriateness of the answers to questions dealing with the main targets of anxiety and worry, and found that the main themes were physical symptoms and self-protection methods. We selected sample questions and answers that addressed these two subjects. Two medical doctors, who are family medicine specialists and the authors of this paper, classified the answers into 5 independent categories: appropriate answers, unrelated answers, wrong answers, advertisement, and other. If there was a disagreement on an answer's category, they discussed until they reached an agreement and recorded the agreed result.

The sample questions were selected by considering the proportion of topics in each question that resulted from using the STM. The themes of physical symptom and self-protection correlated with the fourth and fifth topic communities, respectively. Choosing high proportion questions from each community of topics could form good sample questions that appropriately represent each theme. Therefore, the proportion of the topic communities 4 and 5 was calculated for each question based on the topic proportion information per question. It was created by adding up the proportion of topics belonging to each topic community. The top 100 questions were selected for both topic communities 4 and 5, and the answers to each question were identified. The number of answers was 250 and 306, respectively.

Morphological Analysis and Part of Speech Tagging

For the previously mentioned methods, especially STM and network analysis, to function properly, morphemes need to be extracted from our language data. That is, we need information about what kinds of words appeared from where and at what frequency. The program used to determine this is called a morphological analyzer. From the many types of morphological analyzers applicable to Korean, Komoran was used because it is resilient to the spacing problem in Korean and is sufficiently qualified, as it won awards from the National Institute of Korean Language in 2016.

We would like to mention in advance a caveat regarding the words extracted from the data by Komoran; Korean and English do not respond 1:1. For example, there are several Korean words that can be translated as fever ("열" and "발열"). When translating such words into English, we numbered them (eg, fever_1, fever_2).

Results

Frequency of Documents and Words

Our data included a total of 13,148 questions. Figure 3 presents the number of questions sorted by date.

Of the words that appeared in the questions, Table 1 presents the top 30 frequency words. Those occupying the top 5 positions include "cough," "symptom," "throat," "mask," and "confirmed diagnosis."

Figure 3. Number of questions by date.

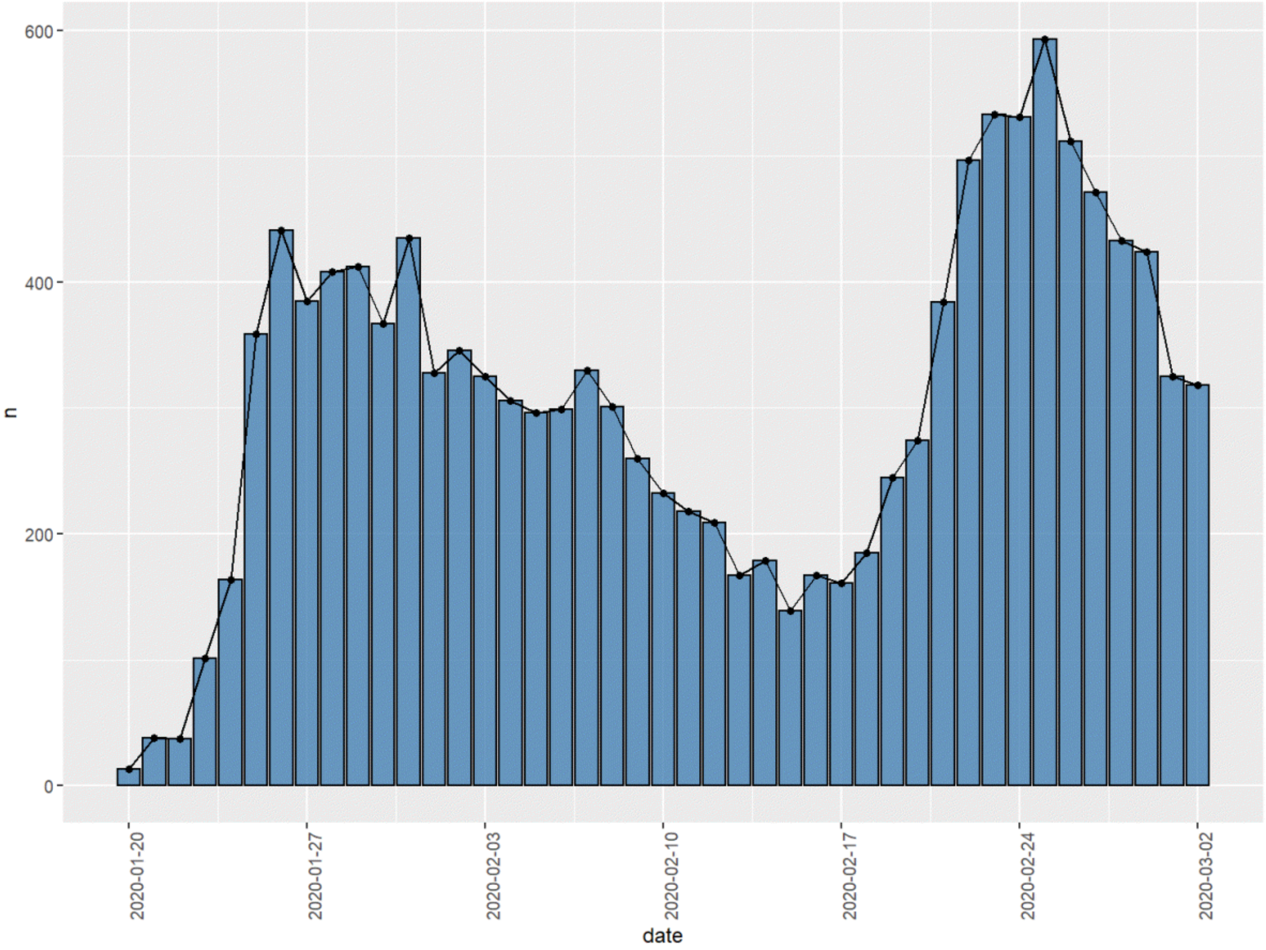


Table 1. Top 30 frequency words in all questions (nouns, excluding words containing “corona”).

Words	Frequency, n
Cough	5589
Symptom	4739
Throat	4177
Mask	2822
Confirmed diagnosis	2199
Human	2095
Cold	1983
Sputum	1953
Fever_1	1907
Degree	1887
Pneumonia	1844
House	1795
Worry	1792
Wuhan	1787
Hospital	1780
China	1691
Rhinorrhea	1432
Head	1427
Infection	1359
Work	1352
Recent	1278
Novel	1268
Headache	1246
Nose	1153
Feeling	1144
Test	973
Travel	958
Body	911
Anxiety	863
Virus	747

Structural Topic Model

About 50 topics were estimated from 13,136 questions using the STM and were interpreted by the authors. Table 2 presents the results of the interpretation. The first column from the left is the topic number, the second column is the topic’s interpretation, and the third column is the topic community number that each topic belongs to. Topic numbers and topic community numbers are nominal numbers for distinction.

Most topics had an apparent subject or consistent contents that allowed a precise interpretation. Some topics, however, were extracted because of expressions that appeared repeatedly in various questions, regardless of the content of the question. This is because topic modeling captures the coappearance patterns of multiple words observed in documents without considering the meanings of words. For example, if many questions on different subjects contain similar expressions, like “please answer my question or you could be cursed,” the topic model would capture that pattern and estimate the topic based on the pattern. In our model, several topics were extracted because of unique Korean language usages. In this case, we unified the interpretation of the topic as “Questions involving particular Korean language expressions without a common subject.” Additionally, we presented the most prominent Korean expression in these topics in Korean in parentheses and translated its meaning into English.

The third column in Table 2 resulted from applying Walktrap to the topics’ correlation network for grouping topics into several cohesive communities. A total of 50 topics were categorized into 6 topic communities, each given a number (the rightmost column). The topics in Table 2 were sorted according to the topic community numbers to help readers easily identify topics belonging to each topic community. Figure 4 presents a visualization of the entire topics network and communities of topics identified through Walktrap. A node is a topic, the number below a node is the topic number, a link is a positive correlation between topics, and the color of the node indicates the topic grouping. Therefore, topics of the same color belong to the same community. Table 3 presents the result of the authors' interpretations of topic communities.

The STM calculates the proportions of topics on a per-document basis. This allowed us to calculate the proportion of each topic in the entire document, which could also be used to calculate the proportion of the topic communities in the entire document. We aggregated proportions of topics belonging to the same community. Additionally, the STM calculated the proportion variation of all topics over time because the time variable was set as a covariate in our STM. The estimates of topics' proportion changes over time could also be aggregated to produce the proportion change of the topic communities over time. Figures 5 and 6 present the results.

Topic community 4 (questions suspecting possible COVID-19 infection after developing a particular symptom) occupied the largest proportion of all the questions. It increased sharply in late February when the number of infections had begun to increase earnestly in Korea. In Figure 6, the yellow line (topic community 4) and the dashed line (the number of confirmed patients) are almost parallel after the second increase for topic community 4, the yellow line mimicking the increase in the number of confirmed patients. Moreover, it is noticeable that the portion of topic community 2 (concerns over working conditions caused by COVID-19) increased slightly as the COVID-19 situation became prolonged.

Table 2. Interpretations of topics.

Topic number	Interpretation	Topic community number
14	Questions on interpretation of English text related to COVID-19[a]	1
21	Questions on daily life that have little relevance to COVID-19	1
34	Questions on using public transport	1
37	Questions suspecting possible COVID-19 infection from past activity or experience	1
39	Questions on screening centers or hospital visits	1
44	Questions suspecting or worrying about possible COVID-19 infection by going outside or coming in contact with people	1
45	Questions suspecting or worrying about possible COVID-19 infection in their family	1
49	Questions involving particular Korean language expressions (“~때”: a noun indicating the particular point in time[b]; “~문제”: problem regarding ~) without a common subject	1
28	Questions regarding the impact of COVID-19 on their working conditions or days off work	2
30	Questions regarding part-time workers coping with the difficulties from COVID-19	2
1	Questions on the current status of the COVID-19 outbreak in Korea	3
2	Questions on the current status of the COVID-19 outbreak outside of Korea	3
3	Other questions regarding COVID-19	3
5	Questions on the expectation of further spread of the COVID-19 outbreak	3
6	Questions on the specific information of confirmed patients in a regional area, including patients’ “route map”	3
10	Questions regarding the end of the COVID-19 epidemic.	3
12	Questions involving particular Korean language expressions (“~가궁금하다”: I wonder~; “~를알려달라”: please explain about~) without a common subject	3
13	Questions expressing the author’s fear and difficulty related to COVID-19	3
15	Questions about treatment of COVID-19 in hospitals	3
18	Questions on specific circumstances that lead to COVID-19 infection	3
23	Questions involving particular Korean language expressions (“부탁드립니다”: could I ask~ please) without a common subject	3
24	Questions on the impact of the COVID-19 outbreak on the stock market	3
25	Questions relating to religion and donation	3
26	Questions about government policies on COVID-19, including an entry ban	3
29	Questions related to the Sincheonji Church and religious events	3
32	Questions involving particular Korean language expressions (“가능할까요”: is it possible to~) without a common subject	3
33	Questions involving particular Korean language expressions (“완전”: an adverb expressing the importance, such as completely or totally; “~겁니다”: an auxiliary verb indicating the current status[c]) without a common subject	3
35	Questions on a COVID-19 vaccine and treatment	3
40	Questions about Wuhan pneumonia in China	3
46	Questions on school opening and delay of school opening; questions on postponing the beginning of the school semesters and reopening date	3
47	Questions involving particular Korean language expressions (“~한데”: I feel like~ then; “~되는데”: it comes to~) without a common subject	3
50	Questions involving particular Korean language expressions (“요즘”: these days or at the moment) without a common subject	3
4	Questions suspecting possible COVID-19 infection after developing a sore throat	4
9	Questions suspecting possible COVID-19 infection after developing chest discomfort	4

Topic number	Interpretation	Topic community number
16	Questions suspecting possible COVID-19 infection after developing a febrile sense and respiratory abnormality	4
19	Questions suspecting possible COVID-19 infection after developing a fever	4
22	Questions on the risk of catching COVID-19 from an exposure to cold weather or in day-to-day living	4
27	Questions suspecting possible COVID-19 infection after developing respiratory discomfort	4
31	Questions suspecting possible COVID-19 infection after developing a headache and cough	4
38	Questions suspecting possible COVID-19 infection after developing diarrhea and eating particular food	4
43	Questions suspecting possible COVID-19 infection after travelling	4
8	Questions about personal protection such as wearing a mask	5
11	Questions on the novel coronavirus contamination mechanism	5
17	Questions about how to use a mask	5
41	Questions on nonscientific knowledge related to COVID-19	5
42	Questions about hand sanitizer and parcels	5
48	Questions about disinfectants and air purifiers	5
7	Questions about the safety of traveling or visiting public places	6
20	Questions about safety related to travel	6
36	Questions on travel and flight cancellation procedures and fees	6

[a] COVID-19: coronavirus disease.

[b] Similar to *when* in a sentence (ie, *when* I was young).

[c] Similar to *have* in a sentence (ie, I *have* a headache).

Figure 4. Topics network and topic communities.

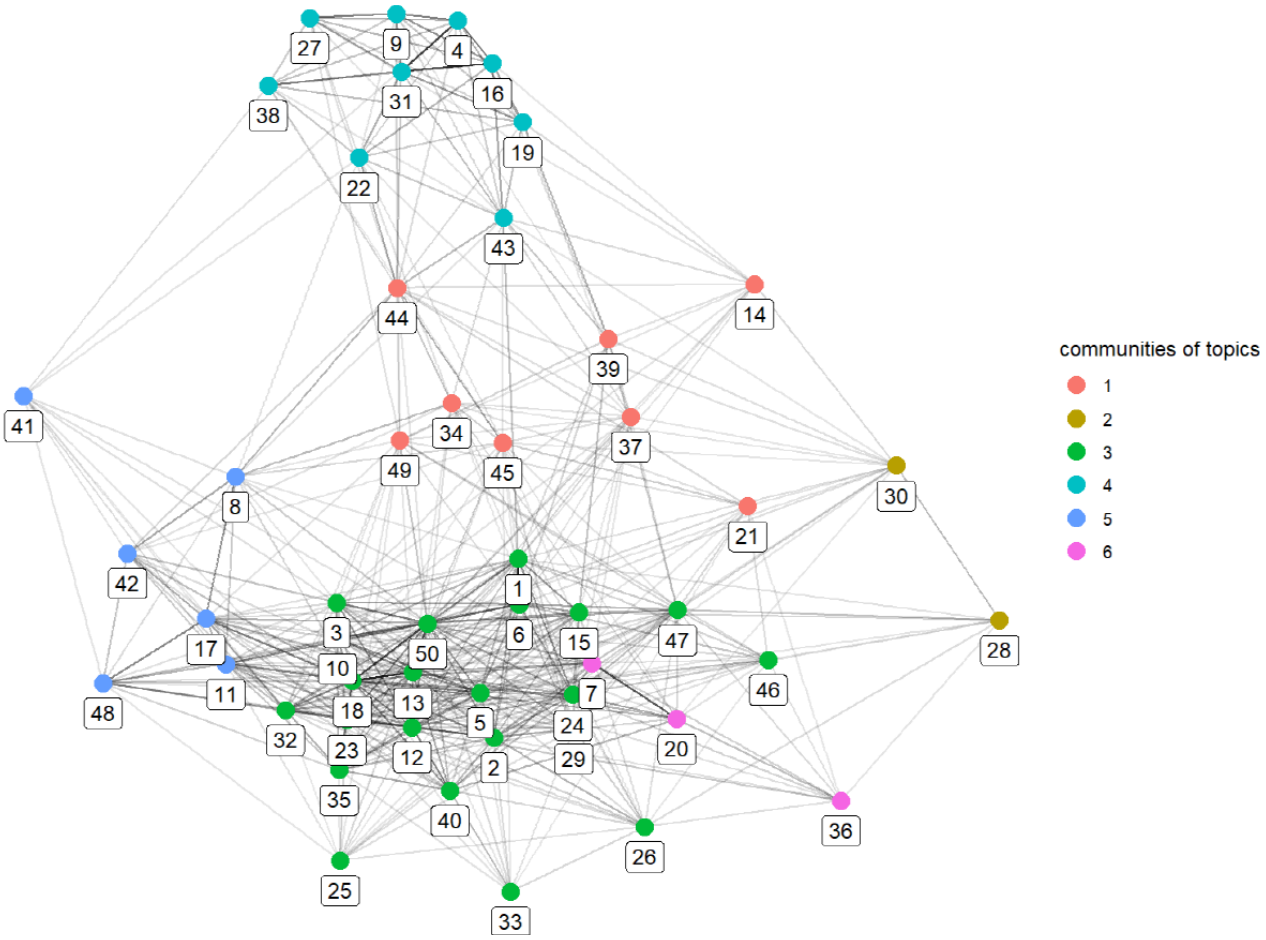


Table 3. Interpretations of topic communities.

Topic community	Interpretation
1	General concern about COVID-19[a] infection
2	Concerns over working conditions caused by COVID-19
3	Questions about the current status of COVID-19 in Korea and government policies
4	Questions suspecting possible COVID-19 infection after developing a particular symptom
5	Questions related to self-prevention measures such as wearing a mask
6	Questions about travel and going out

[a] COVID-19: coronavirus disease.

Figure 5. Proportion of topic communities in all questions. COVID-19: coronavirus disease.

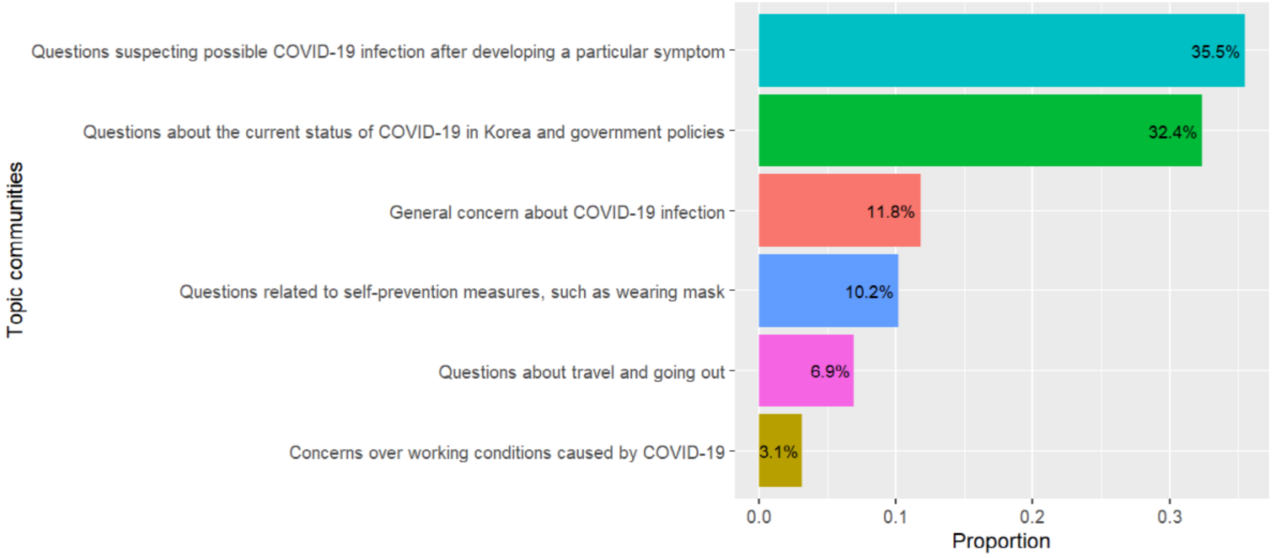
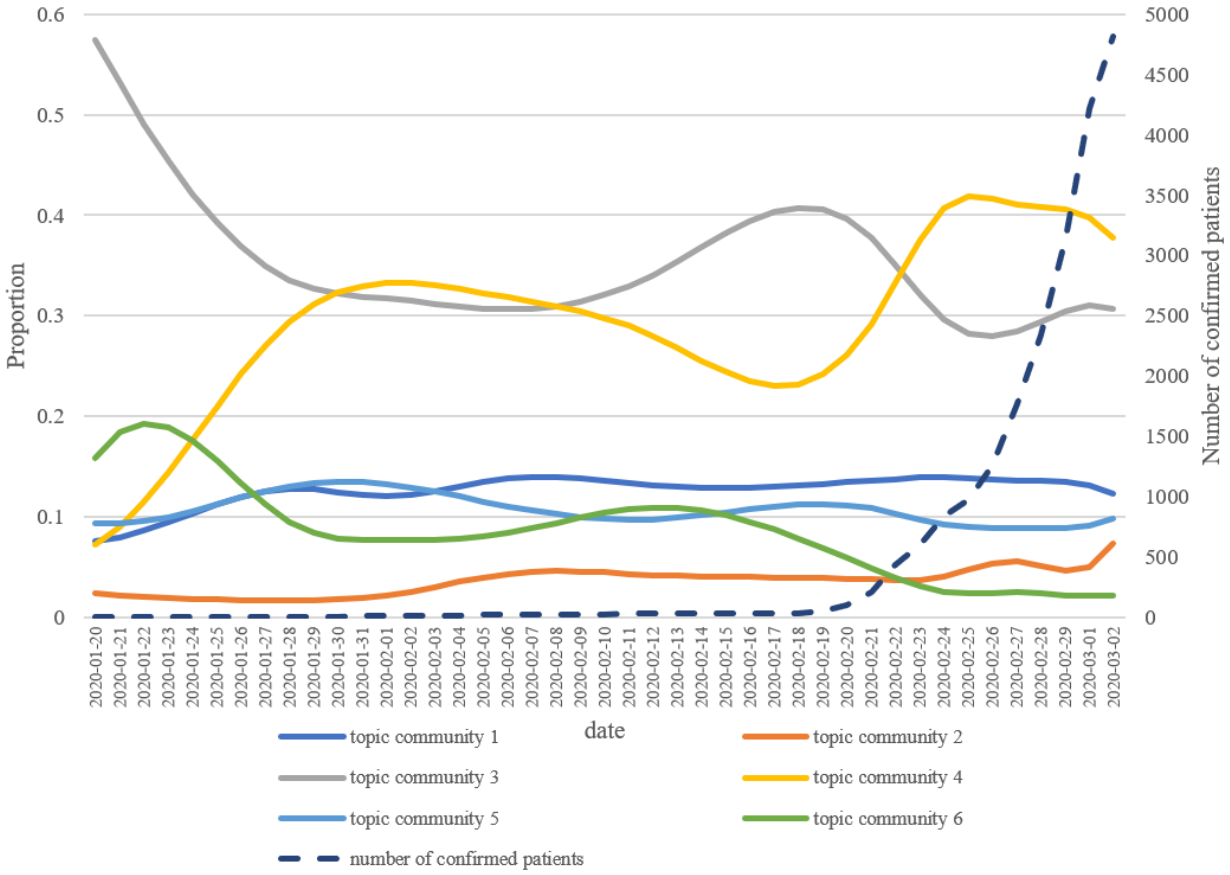


Figure 6. Proportion of topic communities by date.



Network Analysis

We extracted 20 words that were most frequently connected with words referring to anxiety and worry (in Korean, "불안" and "걱정," respectively) from the questions posted from January 20 to March 1, 2020. Table 4 presents the results. In Table 4, the top 5 words include "cough," "symptom," "throat," "mask," and "cold." This allowed us to infer that people's anxiety was centered on physical symptoms and key self-protection methods, such as wearing masks.

We also checked whether the list of words associated with anxiety and worry would vary with time. The period (January 20-March 1, 2020) was divided into 6 weeks, and the top words associated with anxiety and worry was extracted from the data of the 6 subperiods (Figure 7). During the week of January 20-26, when the first confirmed COVID-19 cases were reported, the primary subject of anxiety was “China and traveling.” This was natural, considering that COVID-19 was limited to the Chinese Mainland at that time. However, as the number of confirmed COVID-19 cases increased in Korea, words related to physical symptoms appeared as the top words. The period after February 17 was when the number of confirmed cases increased sharply in South Korea by 2 or 3 digits. Henceforth, Koreans needed to be careful about contact with confirmed patients, and the word “confirmed diagnosis” emerged as the main word associated with anxiety. The word “mask” was consistently included in the top 10 words in all periods. In other words, the anxiety and worry about self-protection have been prevalent consistently throughout the entire period.

We formed a word network using the top 50 words linked to anxiety and worry and extracted its subnetwork based on the most prominent 500 links. By applying the Walktrap algorithm to the subnetwork, we could extract three cohesive word communities or three distinct themes. Figure 8 presents a visualization of the three sets of words comprising at least 5 words. One is related to physical symptoms, another is related to self-protection, and the last is related to China. In other words, people’s anxiety was expressed in three main themes. This result was consistent with our reasoning based on the types of words linked to anxiety and worry.

Table 4. Top 20 frequency words appearing with anxiety and worry in all questions (nouns, excluding words containing “corona”).

Word	Frequency, n
Cough	895
Symptom	750
Throat	602
Mask	507
Cold	418
Degree	409
House	404
Confirmed diagnosis	400
Recent	399
Human	386
Hospital	386
Sputum	383
Fever_1	348
Wuhan	322
China	310
Pneumonia	308
Work	274
Rhinorrhea	271
Headache	245
Thought	238

Figure 7. Top 20 frequency words appearing with anxiety and worry in all questions by week (nouns, excluding words containing "corona").

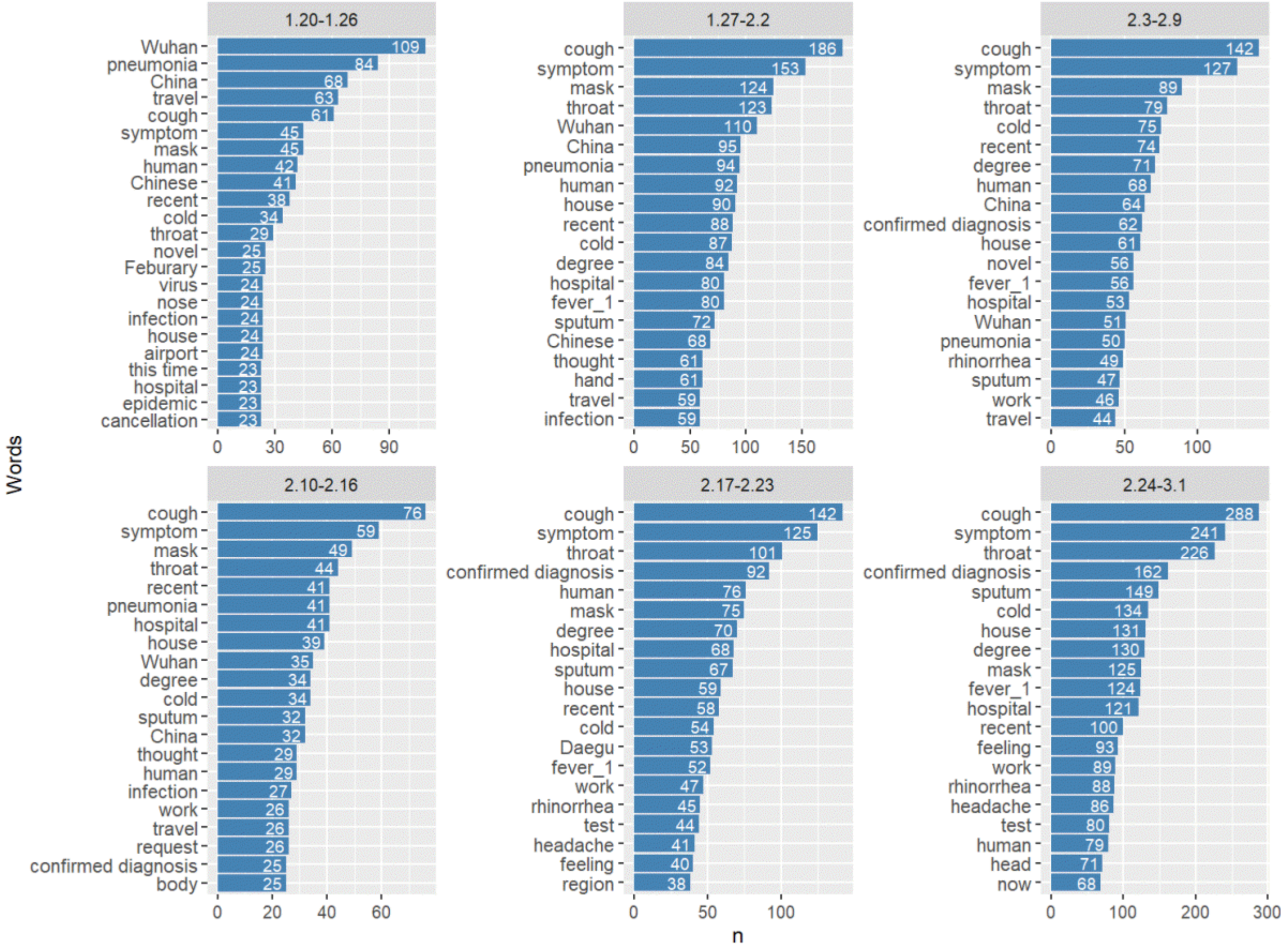
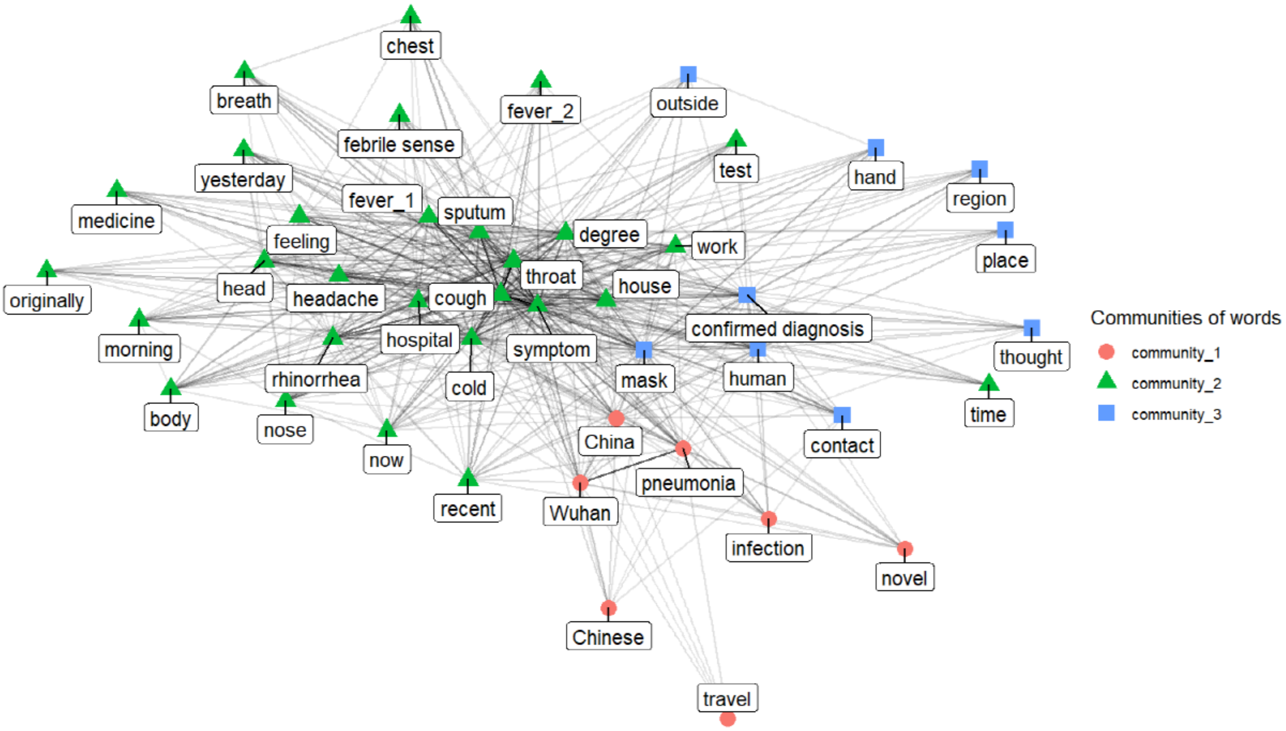


Figure 8. Word network of top 50 words linked to anxiety and worry.

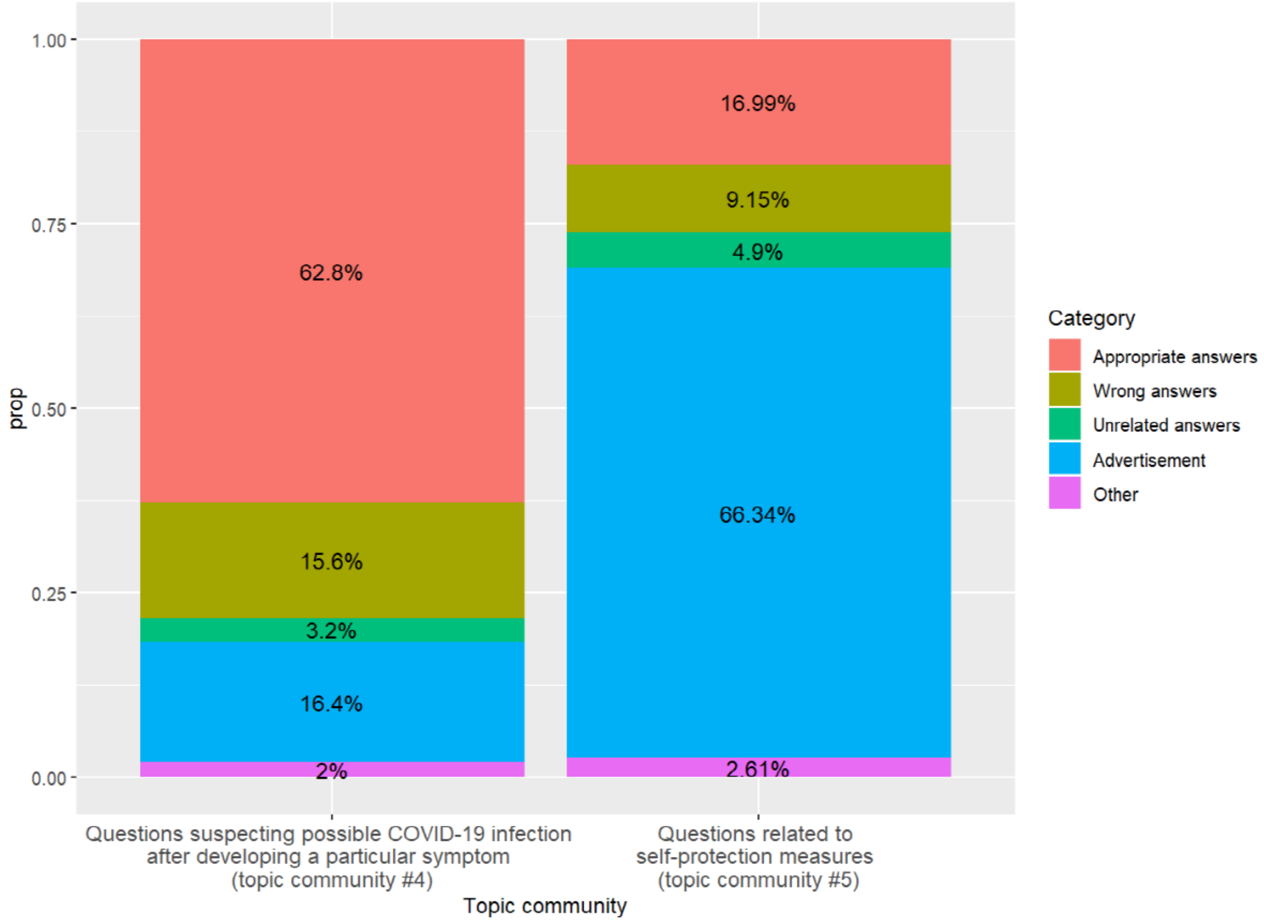


Professionals' Qualitative Coding

Figure 9 shows the results of two medical doctors' categorization of the answers to sample questions dealing with the main targets of anxiety and worry into five categories.

The answers to questions about physical symptoms were often appropriate and relatively less distorted. On the other hand, there were many advertising answers to questions related to self-protection measures.

Figure 9. Proportion of answer categories (based on sample data). COVID-19: coronavirus disease; prop: proportion.



Discussion

In the event of a novel infectious disease outbreak, the general population cannot easily assess the accuracy of the information regarding the disease, and there is increased reliance on online information. Obtaining appropriate and accurate information is extremely difficult, especially in the early stage of an outbreak, due to the uncertainty about the disease. There may be a delay before the governance body, such as health authorities, announces official statements regarding the disease, including symptoms, treatment, or preventive measures. However, it is unlikely that the public would refrain from seeking information and patiently wait until the accurate information becomes publicly available. In this scenario, it is more probable that the members of the populace would seek to acquire the much-needed information from other sources such as the internet.

Consistent with previous studies, our data showed that people sought information regarding COVID-19 on the web. The analysis of 13,136 questions revealed that the largest proportion of topics was regarding anxiety and worries about possible COVID-19 infection after developing a particular symptom. The proportion of topics regarding concerns over working conditions also slightly increased as the COVID-19 outbreak became prolonged. Physical symptoms such as cough, throat pain, and sputum as well as self-protective measures such as wearing a mask were some top key words that simultaneously appeared with the words *anxiety* and *worry* in the word network analysis. This implies that the people were mainly concerned about whether developing a particular physical symptom was relevant to COVID-19 and ways to protect them from COVID-19.

We also analyzed the appropriateness of the answers that replied to the questions. About 63% of the total answers to questions on suspecting possible COVID-19 infection after developing particular symptoms were evaluated as appropriate, while 15.6% of the answers were incorrect, implying a potential dissemination of misinformation. For questions regarding self-protective measures, such as questions asking how to wear masks properly, as much as 66.3% of the answers were advertisements. Thus, it can be assumed that the general population may have difficulty in obtaining appropriate information on self-protective measures.

This study contributes to the establishment of early health communication about public concerns and anxiety observed at the early stage of COVID-19. The initial stage of the epidemic is when neither the health authorities' policies nor people's understanding of the epidemic has stabilized. Under these circumstances, people's online exchange of information and emotion can have a great influence. Governments should implement proper measures to establish online health communication in the early stage of an outbreak to provide appropriate and accurate information. Considering that web data-based studies related to COVID-19 are rare at present, our research's policy and academic value is more pronounced. Although there are studies that use web data to analyze various characteristics of public psychology for other infectious diseases, there are few studies related to COVID-19. Given that numerous countries have been affected recently and are in the initial stage of the COVID-19 outbreak, our study on South Korea, which experienced COVID-19 relatively early, could be a reference point for policy making in other countries.

This study also contributes to devising methods of measuring public psychology using language data in the circumstance of an infectious disease outbreak. Understanding public psychology and culture is essential when dealing with infectious diseases [28] because public psychology greatly impacts the management of infectious diseases [29]. Public anger toward the infected, for example, is common in infectious disease outbreaks, which can induce the infected to hide from quarantine efforts such as screening tests. This is because the infected people will try to avoid the intense social anger directed at them. To prevent this, it is necessary to promptly measure public sentiment in detail and organize appropriate responses such as creating social support for the infected. However, as previously indicated, traditional surveys are relatively difficult to implement quickly and are not free from limitations (eg, effect of researchers' frame on answers, relative difficulty of collecting real time data) [30]. Recently, analysis of language data using computers and statistical models has been introduced, and several scholars have suggested its usefulness [13,31,32]. Our paper provides examples of its use concerning an infectious disease outbreak. Beyond the Q&A forum, various language materials are available on the internet, which can be used to actively supplement existing methods of investigation and create diverse methods to approximate public anxiety. In short, this study shows the potential for "online data-based health policy decision making."

Additionally, this study is differentiated from other studies in how we used text mining techniques. Previously, papers using health-related text data often used frequency as the main information, such as the number of Twitter mentions [33], or applied LDA [34,35], which is the most commonly used topic modeling technique [36]. By using the STM, this study more systematically analyzed how the proportions of COVID-19–related topics varied over time while maintaining the advantages of topic modeling methods. Moreover, this study does not merely apply existing text mining techniques to health-related data but also contributes methodologically on how to use topic modeling methods. Various topic modeling methods have been used as the LDA in many studies [36], most of which draw conclusions from the estimated topics. However, solely focusing on topics has one downside: if the amount of data is substantial, the number of topics will also increase, making it difficult to identify overall patterns that appear in the entire data set. Too many topics, for example 300 topics, represent vast information, which would likely pose a formidable barrier to human researchers. Our research has introduced a way to find sets of topics connected to each other by forming a network of topics and finding cohesive topic communities in them through a network community detection algorithm. In our results, 6 topic communities were found and each community contained content-related topics. It is surprising that the topics belonging to each topic community are meaningfully interrelated, as the topic communities are derived from the community detection algorithm and a correlation matrix deduced by the STM results, not through human researchers' categorization based on the topics' contents. In other words, this study proposes a data-driven method of making topic clusters, which could be used for detecting broader themes from numerous topics that were estimated from a vast number of documents.

Various advanced analyses are possible using the data and results of this study. One of the notable results in our study is that the proportion of topic community 4 (questions suspecting possible COVID-19 infection after developing a particular symptom) seems to be linked to the actual number of confirmed patients. Considering that the symptoms observed in topic community 4 (eg, cough, throat pain, and sputum) were reported as common symptoms of COVID-19 in other clinical reports, this linkage might not be a coincidence. This suggests that information extracted from web data may help identify and even predict the actual trend of infectious diseases. Future research could use sophisticated time series analyses to scrutinize this possibility. Additionally, it is noteworthy that a certain kind of question could have a high proportion of answers written from commercial motivations (Figure 9). This implies that there are many attempts to commercially exploit the infectious disease crisis, and the social effects related to these attempts could be explored using this study's results. Finally, if adequate data of a longer period is available, analyzing how the importance of online information and communication changes over time would be a valuable research project. These further studies will enhance the efficient use of online data for public health.

This study has limitations concerning the range of data. Although Naver is the most popular portal website in Korea, people are not just using this one service. For further comprehensive analysis of information and emotion exchange in online spaces, various sources of web data, including various social network services, need to be incorporated. Furthermore, since the internet user population does not appropriately represent the entire population, it is necessary to consider using the data produced through traditional methods such as surveys and the internet natural language data together.

Nevertheless, this study showed how health information exchanged based on the disease transmission related to people's anxiety and commercial interest in new infectious disease outbreak via a novel approach using online data analysis and topic modeling.

Acknowledgments

This study was supported by research grant from the National Cancer Center, Republic of Korea (1910200-1). This study is not funded by any commercial company, including Naver.com.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2020 Apr 01. Coronavirus disease 2019 (COVID-19) situation report – 72 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn=3dd8971b_2
2. World Health Organization. 2020 Mar 07. WHO statement on cases of COVID-19 surpassing 100,000 URL: https://www.who.int/news-room/detail/07-03-2020-who-statement-on-cases-of-covid-19-surpassing-100-000
3. Rizo CA, Lupea D, Baybourdy H, Anderson M, Closson T, Jadad AR. What Internet services would patients like from hospitals during an epidemic? Lessons from the SARS outbreak in Toronto. J Med Internet Res 2005 Aug 03;7(4):e46 [FREE Full text] [doi: 10.2196/jmir.7.4.e46] [Medline: 16236698]
4. Gui X, Kou Y, Pine K, Chen Y. Managing uncertainty: using social media for risk assessment during a public health crisis. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2017 May Presented at: CHI Conference on Human Factors in Computing Systems; May 2017; Denver, CO p. 4520-4533. [doi: 10.1145/3025453.3025891]
5. Wang W, Ma Y, Wu T, Dai Y, Chen X, Braunstein LA. Containing misinformation spreading in temporal social networks. Chaos 2019 Dec;29(12):123131. [doi: 10.1063/1.5114853] [Medline: 31893637]
6. Auter PJ, Douai A, Makady H, West C. Circulating health rumors in the 'Arab World': a 12-month content analysis of news stories and reader commentary about Middle East Respiratory Syndrome from two Middle Eastern news outlets. Int Commun Gazette 2016 Jul 27;78(5):411-431. [doi: 10.1177/1748048516640202]
7. Barros JM, Duggan J, Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (infoveillance): systematic review. J Med Internet Res 2020 Mar 13;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]
8. Polgreen P, Chen Y, Pennock D, Nelson F. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 01;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
9. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One 2013;8(5):e64323 [FREE Full text] [doi: 10.1371/journal.pone.0064323] [Medline: 23750192]
10. NAVER. 2019. 2018 NAVER annual report URL: https://www.navercorp.com/navercorp_/ir/annualReport/2019/NAVER_AR_2018_Eng.pdf
11. Seo YK, Lee SK. Determinants for the customers' satisfaction of the portal services. e-Business Stud 2010 Mar;11(1):45-68. [doi: 10.15719/geba.11.1.201003.45]
12. OpenSurvey, 2020. 2020. Report on social media and search portal servcies 2020 URL: https://www.opensurvey.co.kr/
13. Kozlowski AC, Taddy M, Evans JA. The geometry of culture: analyzing the meanings of class through word embeddings. Am Sociol Rev 2019 Sep 25;84(5):905-949. [doi: 10.1177/0003122419877135]
14. Hallin DC, Brandt M, Briggs CL. Biomedicalization and the public sphere: newspaper coverage of health and medicine, 1960s-2000s. Soc Sci Med 2013 Nov;96:121-128. [doi: 10.1016/j.socscimed.2013.07.030] [Medline: 24034959]
15. Pitts V. Illness and Internet empowerment: writing and reading breast cancer in cyberspace. Health (London) 2004 Jan;8(1):33-59. [doi: 10.1177/1363459304038794] [Medline: 15018717]
16. Gooden RJ, Winefield HR. Breast and prostate cancer online discussion boards: a thematic analysis of gender differences and similarities. J Health Psychol 2007 Jan;12(1):103-114. [doi: 10.1177/1359105307071744] [Medline: 17158844]
17. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
18. Blei D, Lafferty J. Topic models. In: Text Mining: Classification, Clustering, and Applications. Boca Raton, FL: Chapman and Hall/CRC; 2009:101-124.
19. Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci U S A 2004 Apr 06;101 Suppl 1:5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]
20. Roberts ME, Stewart BM, Tingley D. stm: R package for structural topic models. J Stat Soft 2019;91(2):1-40. [doi: 10.18637/jss.v091.i02]
21. Roberts ME, Stewart BM, Airoldi EM. A model of text for experimentation in the social sciences. J Am Stat Assoc 2016 Oct 18;111(515):988-1003. [doi: 10.1080/01621459.2016.1141684]
22. Wallach H, Murray I, Salakhutdinov R, Mimno D. Evaluation methods for topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning. 2009 Jun Presented at: 26th Annual International Conference on Machine Learning; June 2009; Montreal, Quebec. [doi: 10.1145/1553374.1553515]
23. Fortunato S. Community detection in graphs. Phys Rep 2010 Feb;486(3-5):75-174. [doi: 10.1016/j.physrep.2009.11.002]

24. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge, United Kingdom: Cambridge University Press; 1994.
25. Fortunato S, Hric D. Community detection in networks: a user guide. Phys Rep 2016 Nov;659:1-44. [doi: 10.1016/j.physrep.2016.09.002]
26. Yang Z, Algesheimer R, Tessone CJ. A comparative analysis of community detection algorithms on artificial networks. Sci Rep 2016 Aug 1;6(1):1-16. [doi: 10.1038/srep30750]
27. Blei DM, Lafferty JD. A correlated topic model of Science. Ann Appl Stat 2007 Jun;1(1):17-35. [doi: 10.1214/07-AOAS114]
28. Ogoina D. Behavioural and emotional responses to the 2014 Ebola outbreak in Nigeria: a narrative review. Int Health 2016 Jan;8(1):5-12. [doi: 10.1093/inthealth/ihv065] [Medline: 26678568]
29. Rosenbaum L. Communicating uncertainty — Ebola, public health, and the scientific process. N Engl J Med 2015 Jan 01;372(1):7-9. [doi: 10.1056/nejmp1413816]
30. Babbie ER. The Practice of Social Research. Scarborough, ON: Nelson Education; 2015.
31. Evans JA, Aceves P. Machine translation: mining text for social theory. Annu Rev Sociol 2016 Jul 30;42(1):21-50. [doi: 10.1146/annurev-soc-081715-074206]
32. Jo W. Possibility of discourse analysis using topic modeling. J Asian Sociol 2019;48(3):321-342.
33. Rodriguez-Morales AJ, Castañeda-Hernández DM, McGregor A. What makes people talk about Ebola on social media? A retrospective analysis of Twitter use. Travel Med Infect Dis 2015;13(1):100-101. [doi: 10.1016/j.tmaid.2014.11.004] [Medline: 25468077]
34. Fu K, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702. [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
35. Choi S, Lee J, Kang M, Min H, Chang Y, Yoon S. Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks. Methods 2017 Oct 01;129:50-59 [FREE Full text] [doi: 10.1016/j.ymeth.2017.07.027] [Medline: 28813689]
36. Gerlach M, Peixoto TP, Altmann EG. A network approach to topic models. Sci Adv 2018 Jul;4(7):eaaq1360 [FREE Full text] [doi: 10.1126/sciadv.aaq1360] [Medline: 30035215]

Abbreviations

COVID-19: coronavirus disease
FREX: frequency and exclusivity
LDA: latent Dirichlet allocation
Q&A: questions and answers
STM: structural topic model

Edited by G Eysenbach; submitted 18.04.20; peer-reviewed by S Jung, YC Chen; comments to author 11.05.20; revised version received 26.05.20; accepted 27.05.20; published 02.06.20.

Please cite as:
Jo W, Lee J, Park J, Kim Y
Online Information Exchange and Anxiety Spread in the Early Stage of the Novel Coronavirus (COVID-19) Outbreak in South Korea: Structural Topic Model and Network Analysis
J Med Internet Res 2020;22(6):e19455
URL: https://www.jmir.org/2020/6/e19455
doi:10.2196/19455
PMID:32463367

Viewpoint

Considerations for an Individual-Level Population Notification System for Pandemic Response: A Review and Prototype

Mohammad Nazmus Sakib[1], MBBS, MSc; Zahid A Butt[1], MBBS, MSc, PhD; Plinio Pelegrini Morita[1,2,3,4,5], MSc, PhD, PEng; Mark Oremus[1], PhD; Geoffrey T Fong[1,6,7], PhD, FRSC, FCAHS; Peter A Hall[1], PhD

[1]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada
[2]Research Institute for Aging, University of Waterloo and Schlegel Villages, Waterloo, ON, Canada
[3]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada
[4]eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
[5]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[6]Department of Psychology, University of Waterloo, Waterloo, ON, Canada
[7]Ontario Institute for Cancer Research, Toronto, ON, Canada

Corresponding Author:
Peter A Hall, PhD
School of Public Health and Health Systems
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 38110
Email: pahall@uwaterloo.ca

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/6/e21634/

Abstract

The outbreak of the coronavirus disease (COVID-19), caused by severe acute respiratory syndrome coronavirus 2, spread worldwide after its emergence in China. Whether rich or poor, all nations are struggling to cope with this new global health crisis. The speed of the threat's emergence and the quick response required from public health authorities and the public itself makes evident the need for a major reform in pandemic surveillance and notification systems. The development and implementation of a graded, individual-level pandemic notification system could be an effective tool to combat future threats of epidemics. This paper describes a prototype model of such a notification system and its potential advantages and challenges for implementation. Similar to other emergency alerts, this system would include a number of threat levels (level 1-5) with a higher level indicating increasing severity and intensity of safety measures (eg, level 1: general hygiene, level 2: enhanced hygiene, level 3: physical distancing, level 4: shelter in place, and level 5: lockdown). The notifications would be transmitted to cellular devices via text message (for lower threat levels) or push notification (for higher threat levels). The notification system would allow the public to be informed about the threat level in real time and act accordingly in an organized manner. New Zealand and the United Kingdom have recently launched similar alert systems designed to coordinate the ongoing COVID-19 pandemic response more efficiently. Implementing such a system, however, faces multiple challenges. Extensive preparation and coordination among all levels of government and relevant sectors are required. Additionally, such systems may be effective primarily in countries where there exists at least moderate trust in government. Advance and ongoing public education about the nature of the system and its steps would be an essential part of the system, such that all members of the public understand the meaning of each step in advance, similar to what has been established in systems for other emergency responses. This educational component is of utmost importance to minimize adverse public reaction and unintended consequences. The use of mass media and local communities could be considered where mobile phone penetration is low. The implementation of such a notification system would be more challenging in developing countries for several reasons, including inadequate technology, limited use of data plans, high population density, poverty, mistrust in government, and tendency to ignore or failure to understand the warning messages. Despite the challenges, an individual-level pandemic notification system could provide added benefits by giving an additional route for notification that would be complementary to existing platforms.

(J Med Internet Res 2020;22(6):e19930) doi:10.2196/19930

KEYWORDS

pandemic; epidemic; notification system; hygiene; physical distancing; lockdown; mobile technology; COVID-19; coronavirus

Introduction

Despite a dramatic improvement in health care affordances and extensive public health measures around the world, the emergence and re-emergence of infectious pathogens and associated diseases have become a common phenomenon. The first 2 decades of the 21st century have already witnessed several major disease outbreaks such as severe acute respiratory syndrome (SARS) in 2003, with the most severe outbreak being the coronavirus disease (COVID-19) pandemic [1,2].

In December 2019, the outbreak of COVID-19, caused by a novel coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first reported in Wuhan, China [3,4]. Full-genome sequencing and phylogenic analysis revealed that SARS-CoV-2 is distinct from both severe acute respiratory syndrome–related coronavirus (SARS-CoV) and Middle East respiratory syndrome–related coronavirus (MERS-CoV) but closely linked [5]. Within a brief period, the disease has spread to virtually all regions of the world with a total of 6,112,902 confirmed cases and 369,593 deaths as of May 31, 2020 [6]. The World Health Organization (WHO) announced COVID-19 a “Public Health Emergency of International Concern” on January 30, 2020, and subsequently declared a pandemic on March 11, 2020 [7].

COVID-19 is primarily a respiratory illness and the majority of patients present with flu-like symptoms such as fever, cough, fatigue, shortness of breath, headache, and muscle pain; although, a significant minority experience different symptoms, some of which are highly distinctive (eg, ageusia, anosmia) [1,3,8,9]. Although most patients usually develop mild illnesses (81%), it can cause severe pneumonia, respiratory failure, and even death, particularly in individuals with an existing high-risk condition such as chronic obstructive pulmonary disease or diabetes [1,3,8,9]. It has been reported that about 14% of the patients develop a severe disease that requires hospitalization and oxygen support, and 5% require intensive care support [4,10]. Although the case fatality rate is expected to be around 0.5-1% [11,12], it is significantly higher among older adults and those with a high-risk condition [1,4,8]. Accordingly, countries with a higher older demographic are experiencing more critically ill patients and mortality in absolute terms [13,14].

In comparison to SARS and Middle East respiratory syndrome (MERS), which showed a mortality rate of 10% and 34%, respectively [15,16], COVID-19 appears to be less deadly. However, a sharp initial increase in the number of cases indicates that SARS-CoV-2 could be more contagious than the previously emerged coronaviruses [17]. The basic reproduction number (R_0) of SARS-CoV-2, which indicates the average number of new cases generated from an infected person, is comparable to SARS [18]. Nevertheless, as a large number of patients develop a mild or asymptomatic form of the illness and may not seek medical advice, it greatly increases the risk of community transmission and larger outbreaks [19,20].

Overall, the emergence and re-emergence of highly infectious pathogens pose a substantial threat to human health. Air travel and tourism now allow for epidemics to quickly become pandemics [21]. In fact, the current pandemic was predicted and foreseen previously on multiple occasions [22]. The lessons from the ongoing global health crisis indicate that a substantial improvement is needed in our pandemic response, and this is possible by making use of mobile technology. The impact of future outbreaks can be mitigated through an improvement in pandemic notification systems by harnessing such technology.

Existing Global Platforms for Infectious Disease Surveillance

The Program for Monitoring Emerging Diseases (ProMED; also known as ProMED-mail) is an internet-based, publicly available, and free of charge digital surveillance system dedicated for rapid collection and circulation of information on emerging infectious diseases and toxins that affect humans, animals, and plants [23,24]. This novel platform was founded in 1994 by the Federation of American Scientists and later became a part of the International Society for Infectious Diseases in 1999. ProMED has grown substantially since its establishment, and the number of subscribers increased to approximately 80,000, which represents every country in the world [23,24]. The ProMED system collects information from a variety of sources (eg, media reports, official reports, local observers); a multidisciplinary team of experts from different countries synthesizes the information, which is then disseminated to the subscribers directly by email and posted online on the ProMED website. The ProMED system continuously monitors emerging disease status and delivers updates as required in near real time. This platform was the first to report major disease outbreaks including SARS, MERS, Ebola, and Zika [23,24]. ProMED is now being used by a wide range of professionals and organizations including government officials, private organizations, researchers, physicians, and journalists.

Another global platform to address emerging public health issues is the Global Public Health Intelligence Network (GPHIN), developed by the Government of Canada in collaboration with the WHO [25]. Unlike ProMED, GPHIN is a fee-based service with subscription offers limited to only selected organizations involved in public health issues. GPHIN gathers information from global media sources (ie, Factiva and

Al Bawaba) on a variety of public health topics [26,27], which is then filtered by an automated process for appropriateness and made accessible to the WHO, relevant government authorities, and subscribers. GPHIN reports constitute approximately 40% of the WHO's early warning outbreak information [25]. Several other notable web-based disease surveillance systems are Influenzanet, the Global Outbreak Alert and Response Network, and Google Flu Trends [28-30].

Regarding the COVID-19 outbreak, both the ProMED and GPHIN systems delivered early notifications in late December 2019 of what turned out to be SARS-CoV-2, the virus causing COVID-19 [31]. Subsequently, SARS-CoV-2 spread worldwide, despite the presence of these early surveillance systems. It should be noted that these systems operate mostly at the level of government and public health authorities; therefore, an active engagement of the public throughout the process, which is a requirement for the preventive measures to be of utmost effectiveness, is an ongoing concern. Administration of an individual-level, population-wide alert system would be complementary to the existing platforms by enhancing public response and engagement.

Rationale for an Individual-Level Pandemic Notification System

The COVID-19 pandemic emerged with such rapidity and lack of clarity on its transmissibility and severity that many governments implemented a cascade of changing, ad-hoc actions to address the situation [32,33]. In Canada, SARS and Ebola had primed the health care system to develop refined measures and protocols for dealing with cases, and to protect frontline health care workers [34,35]. Yet, there have been medical supply shortages and ongoing challenges to produce plans for managing incoming patients [36-40]. In the 8-week period leading up to the declaration of lockdown and social distancing measures, Canada, for example, moved from no travel restrictions and a declared low risk to the Canadian public to imposition of travel restrictions but not bans and selective banning of public gatherings [41]. These were then followed quickly by outright travel and gathering bans, coupled with stay-at-home orders [41]. This dynamic response played out across many countries around the world and continues as countries begin to lift restrictions. In the early stages, rapidly changing responses to the unfolding pandemic contributed to public confusion about expectations.

Throughout, the North American public has received communications from many sources of information, including those made by the Public Health Agency of Canada, the US Centers for Disease Control and Prevention, and regional counterparts. As we have witnessed with previous outbreaks, the COVID-19 pandemic has spawned a parallel, massive infodemic, leading to a surge of fabricated news and distorted evidence, which has increased the potential for health hazards while substantially diminishing the impact of valid information issued by health authorities about facts on all aspects of the virus, including origin, transmissibility, severity, and preventive measures [42-46]. Much like earlier trends, the presence of misinformation, rumors, and hoaxes about the novel coronavirus was also omnipresent throughout social media and the internet, which ultimately inflamed the crisis. To illustrate, there were claims about the origins of the virus (ie, bioterrorism and conspiracy theories) that led to xenophobia in many regions of the world [47,48]. Furthermore, some potentially dangerous rumors about miracle remedies spread in social media (ie, consumption of bleach, methanol, and disinfectant), which led to the deaths of people who were otherwise healthy and noninfected [49-51].

In addition, the media's sensationalistic coverage of extreme events and reports without context (eg, bodies on the streets of Ecuador, Italian military hauling coffins) exacerbated the public reaction [52,53]. A spike in fear-driven panic buying of masks and other medical supplies occurred initially [54,55], ultimately contributing to a subsequent shortage on the front lines among hospital staff in some countries [36,37]. This suggests that some of the fear-driven responses among the public may have misdirected behavior in the early stages, in part due to confusion about what constituted an appropriate response to the level of threat at a given point in time.

Ideally, the public should be informed and advised by clear communications in a manner that can rise above the buzz and chaos of the plethora of information sources in the world. It must be evidence-based and driven by public health authorities who have the expertise to understand and interpret the rapidly growing scientific studies on the novel pathogen and can synthesize the material and communicate it in ways that the public can readily understand, and to provide advice in ways that the public can readily follow [43]. This may facilitate rule-based (better than emotion-based) responding among the population and diminish the unintended affect-driven responses that may be not desirable. Considering these matters, an individual-level, graded pandemic notification system, if implemented nationwide, could provide a significant benefit complementary to existing systems by ensuring a timely and coordinated response, transmitting notification directly to the general population via mobile devices. This will likely be important during multi-wave dynamic pandemics such as COVID-19 and during oscillation between response levels (ie, imposing and lifting restrictions).

Prototype Model for an Individual-Level Pandemic Notification System

As COVID-19 has spread worldwide with catastrophic effects on both economy and human life, there is a renewed consideration of an individual-level, general-purpose pandemic notification system responsive to outbreaks of an infectious illness. Such a system would make responses more universally clear and decipherable for the average member of the population. Much like a disaster or weather notification system, a pandemic notification system would be structured as a graded level of threats. Each level would ideally provide details on three aspects of the outbreak: information about the pathogen, level of threat, and behavioral directives (eg, hygiene, physical distancing). This system, known in advance by the public, would be used to prime the public in real time with respect to changes in threat level and, equally importantly, the expected responses. For example, if the system was operational now (Textbox 1, Figure 1), level 1 or 2 would provide information about SARS-CoV-2, the impending threat nationally and locally, and hygiene measures that need to be followed; threat level 3 would advise people toward physical distancing; level 4 would push stay-at-home orders; and level 5 would be more extreme measures such as a lockdown. Overall, the threat levels would be cumulative, and the public would follow all the precautionary measures outlined in the lower threat levels on top of the precautionary measures required by the current threat level put in place. Depending on the circumstances, the threat level could vary from region to region in a given country and could be switched as a given situation required. The mode of delivering the alert would be adjusted based on the intensity of the threat. For instance, the alert for a threat level of 1 or 2 may be disseminated by simple text messaging or passive information observed only for those who look at the relevant app or website source, but the alert for threat levels 3, 4, and 5 could be push notifications directly to the mobile phones of the entire population, similar to an Amber Alert (an Amber Alert is an emergency response system that provides immediate and up-to-date information to the public about a recently missing or abducted person) in Canada and other countries.

Textbox 1. Sample operational definitions and guidelines for a graded pandemic notification system. The threat levels are cumulative. Level 3A and above must include Enhanced Hygiene by default.

Level 5: lockdown

This is the extreme level of precautionary measure, which requires a particular neighborhood to completely shut down all necessary and nonessential outdoor activities by the residents, with the exception of only law enforcement agencies, health care professionals, and personnel involved with essential services. Residents will contact the authority for emergency necessity and can go outside only with authorization. It is warranted when a large disease outbreak is confirmed in close vicinity, and the residents are exposed to severe risk of disease transmission.

Level 4: shelter in place

This is a high level of precautionary measure, which requires a particular neighborhood to completely shut down all nonessential outdoor activities by the residents. People can go outside only for necessary activities, which include but are not limited to buying food, getting medication, and essential in-person doctor’s appointments. It is warranted when a large disease outbreak is confirmed in close vicinity, and the residents are exposed to a high risk of disease transmission.

Level 3C: physical distancing (strict)

This is a high level of precautionary measure, which requires a particular neighborhood to follow strict physical distancing guidelines. Citizens must maintain a specific distance from each other while outdoors, with the exception of family members or individuals living in the same household. Any gathering of more than 2 people is prohibited, and noncompliance is subject to legal sanction. All schools and offices will be closed, and work from home protocol will be activated. It is warranted when a disease outbreak is confirmed in close vicinity, and the residents are exposed to a significant risk of disease transmission.

Level 3B: physical distancing (moderate)

This is a moderate level of precautionary measure, which requires a particular neighborhood to follow moderate physical distancing guidelines. Citizens must avoid large gatherings of a specific number (to be determined by the context of the locality). Offices can be open with caution and maintaining physical distancing guidelines. Schools will be closed except for the necessary activities. Course activities will be continued through online platforms. It is warranted when a disease outbreak is confirmed in close vicinity, and the residents are exposed to a significant risk of disease transmission, but the risk is lower than the Level 3C (to be determined by the number of cases, disease trend, and other vital information of the respective locality and surroundings).

Level 3A: physical distancing (mild)

This is a moderate level of precautionary measure, which requires a particular neighborhood to follow mild physical distancing guidelines. Citizens must avoid large gatherings of a specific number (to be determined by the context of the locality). Offices and schools can remain open with caution and maintaining adequate physical distancing guidelines. It is warranted when a disease outbreak is confirmed in close vicinity, and the residents are exposed to a significant risk of disease transmission, but the risk is lower than the level 3B (to be determined by the number of cases, disease trend, and other vital information of the respective locality and surroundings).

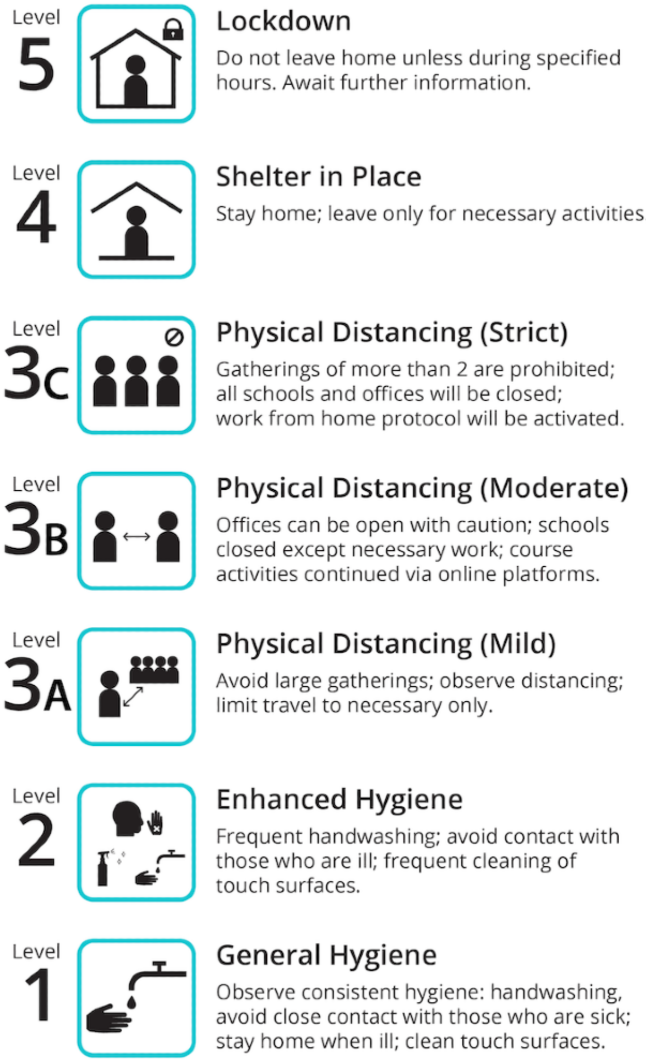
Level 2: enhanced hygiene

This is a mild level of precautionary measure, which requires a particular neighborhood to follow enhanced hygiene guidelines that include but are not limited to frequent handwashing, wearing masks where applicable, avoiding contact with people who are ill, and cleaning touched surfaces. It is warranted when a disease outbreak is suspected in close vicinity, and the residents are exposed to a general risk of disease transmission.

Level 1: general hygiene

This is a low level of precautionary measure, which requires a particular neighborhood to follow general hygiene guidelines that include but are not limited to observing consistent hygiene, handwashing, avoiding contact with people who are sick, staying home when ill, and cleaning touched surfaces. It is warranted when a disease outbreak is suspected in close vicinity, and the residents are exposed to a low risk of disease transmission.

Figure 1. Sample icon and instruction set for a pandemic alert and response notification system.



Lockdown
Do not leave home unless during specified hours. Await further information.

Shelter in Place
Stay home; leave only for necessary activities.

Physical Distancing (Strict)
Gatherings of more than 2 are prohibited; all schools and offices will be closed; work from home protocol will be activated.

Physical Distancing (Moderate)
Offices can be open with caution; schools closed except necessary work; course activities continued via online platforms.

Physical Distancing (Mild)
Avoid large gatherings; observe distancing; limit travel to necessary only.

Enhanced Hygiene
Frequent handwashing; avoid contact with those who are ill; frequent cleaning of touch surfaces.

General Hygiene
Observe consistent hygiene: handwashing, avoid close contact with those who are sick; stay home when ill; clean touch surfaces.

Figure 1 presents an example of a symbol and number-coded system that is easy to remember and interpret by members of the public, even among those with limited literacy. Features include a hierarchical arrangement of threat levels and associated responses, the latter including links to more detailed information as necessary. If this were a universal standard within a given country, health care settings, workplaces, and educational institutions would also have a common understanding with employees, students, parents, and others on a moment-to-moment basis as to what is expected. Similar approaches have been widely tested and implemented in other domains, such as Defense Readiness Condition alert states [56] and travel advisories [57]. The integration of public education efforts along with drills or simulations during the time between threats would be useful to ensure that all have some level of recent familiarity with procedures and aspects of their implementation. Each of these settings could also be encouraged to include a response plan for their own institution, such that institutional protocols are worked out in advance.

If such a platform is available in a country capable of transmitting notification directly to the population through mobile messaging, it would improve the clarity of communication about both the changing threat level (which, in the case of an emerging pandemic, can be day-to-day) and what the expected steps are at each of these threat levels, as well as data-driven tools for helping public health officials to plan and monitor interventions. This could in turn be linked to reliable website information and data visualization for the public. The technology necessary for implementing these solutions is widely available and can easily be used to deliver the proposed services. However, major challenges lie in the privacy and governance of such platforms, requiring the guidance of a team of expert researchers in public health, privacy, and public policy to facilitate the implementation of such a solution.

Expectations on the level of everyday citizens, employers, and health care professionals could be aligned more easily by proactive creation of such a system, coupled with public education as to how to interpret and act on it. This latter facet is critically important—public education about the meaning of each level and the expected sequence of actions to take in response will reduce fear by making the response more obvious when the signal comes.

The public education and infrastructure investment for a pandemic notification system may require considerable funding; however, it could provide a long-term benefit to handle future pandemics once developed. Compared to SARS and MERS, COVID-19 will be costly to some countries due to delayed response and inefficient implementation of the recommendations [58,59]. A notification system would mitigate the initial confusion and improve implementational efficiency, which could, therefore, be economical in the long run.

Prospects of Pandemic Notification Systems in Developed Countries

Developed countries have long-standing experience operating different alert systems successfully, such as the Emergency Alert System, Amber Alert, and Alert Ready. However, existing alert systems have been criticized for being inefficient for a myriad of reasons, such as the execution of false alarms, nonresponse, uneven distribution, network disruption, inadequate technology, and inappropriate timing [60-62]. Despite that, the necessity of emergency notification systems is indisputable, as a substantial number of lives could be saved and casualties could be prevented in the event of natural and human-made disasters [63,64].

Although the COVID-19 pandemic has spread pervasively, numerous strategies have been adopted by developed countries to communicate with the public about the threat and preventive measures undertaken. Some countries have excelled in their battle against COVID-19 because of prompt, coordinated, and organized actions. New Zealand, so far, is leading the battle against COVID-19 [65]. From the beginning, New Zealand has adopted the strategy of eliminating the virus rather than just containing it by imposing a national lockdown and other stringent measures [66]. In addition, a multilevel COVID-19 alert system, similar to terror alert, has been launched by the government of New Zealand to counter the ongoing crisis. This system comprises of 4 levels of alerts based on the impending risks of disease transmission: level 1"Prepare," level 2 "Reduce," level 3 "Restrict," and level 4 "Lockdown" [67]. A higher level indicates an increased risk of disease transmission, therefore, associated with more stringent measures and restrictions. Each level has a risk assessment coupled with a range of public health and social measures expected to be followed by the residents. This is an organized and efficient way to communicate with the public about day-to-day restrictions in effect. New Zealand's strategy appears to be working, as only 1504 confirmed cases and 22 deaths have been reported as of May 31, 2020, [68] and the transmissibility has been reduced to below outbreak threshold [65].

The United Kingdom has recently launched a five-tier coronavirus alert system to handle the pandemic situation more efficiently [69,70]. In this alert system, the threat levels of SARS-CoV-2 have been ranked on a scale of 1-5 with color coding for easy comprehension of the threat for the public: level 1 "Low" (green), level 2 "Moderate," level 3 "Substantial" (orange), level 4 "Severe," and level 5 "Critical" (red) [69,70]. The threat levels range from an absence of the virus corresponding to level 1 to a rapid ongoing transmission corresponding to level 5. A higher threat level indicates tougher social distancing and lockdown measures. The changes in the threat levels will be made based on medical and scientific data, such as the number of cases, the R_0, and transmission rate, and will be operated by the Joint Biosecurity Centre in collaboration with the government. This system is expected to facilitate a relatively smooth and efficient transition between threat levels. When the system launched on May 10, 2020, the United Kingdom was declared to be in level 4 (severe), which denotes partial lockdown measures, ongoing community transmission, and slightly stretched yet coping health care systems [69].

Compared to New Zealand and the UK systems mentioned, the prototype model we propose has several advantages. First, we adopted precautionary measures to define the threat levels rather than the less descriptive terms used in the existing systems. For example, if a threat level is labeled as "severe" or "restrict," it would not be readily understandable by the public unless they thoroughly study the alert system. The use of the terminology "prepare," "reduce," and "restrict" could also confuse people about which levels are being referred to if these are not accompanied by the numerical threat levels. The terms "substantial," "severe," and "critical" could mean the same intensity to the public and, therefore, always need to be followed by the level number when mentioned. In contrast, the use of precautionary measures as a level description is self-explanatory because common people understand the meaning without any additional effort. For example, if we say that the "Strict Physical Distancing" level is in effect, people would be able to visualize at which point they are in the hierarchy of the notification system. Second, the use of one logo per level could have an additional benefit, especially for people who are illiterate, as they can readily comprehend it by observing the logo. Third, as these terms of precautionary measures have become universal because of the current pandemic, any visitor in the foreign county would also be able to understand the prevention measures they must abide by for the current threat level in effect.

Aside from these newly implemented alert systems, sporadic use of the emergency notification systems has also been observed. In the province of Ontario, Canada, emergency alerts have been dispatched to inform dwellers about self-isolation and stay-home orders [71]. In addition, the use of existing disaster alert systems for the purpose of the COVID-19 pandemic was also evident. For example, Alert Bay, a remote coastal British Columbia community, has been using a tsunami warning siren to inform the residents about the COVID-19 curfew in the locality [72]. However, there is still a lack of a coordinated and systematic approach, which hinders the efficiency of the system and the effectiveness of the control measures [73]. This increases the possibility that critical recommendations, modifications, and warnings would be overlooked, especially for those who are not actively looking for the information. It might be a contributing factor that many residents ignored stay-home orders at the beginning [74]. Therefore, only imposing preventive measures would not be conducive unless citizens, for whom such measures are intended, are informed proactively. A recent survey by Calgary-based Public Emergency Alerting Services Inc is worth noting, which reported that about 84% of Canadians believe that the public should be informed about COVID-19 situations through the national public alert system [75]. It signifies the fact that citizens wish to be advised by the official sources in the event of a national crisis, at least in Canada.

Therefore, in similar developed countries where mobile phone penetration is high and wherein the trust of government is at least moderately high, a graded pandemic notification system may be a potential advantage over the status quo. The efficiency of the proposed notification system, however, would largely depend on the adherence and compliance of the citizens to the system, and how promptly the system could be activated. For instance, a measure like physical distancing would be most effective in containing an epidemic only if the public strictly complies with it and commences at the onset of the epidemic [76-79]. An organized, systematic, and coordinated approach is needed to receive the ultimate benefit from the system. The incorporation of legal enforcement could be necessary for the optimal functioning of the system, and the precautionary measures related to the threat levels should be prompt upon notification. To illustrate, when a lower threat level is expected and dispatched (level 1-2), the public would efficiently follow general or enhanced hygiene as instructed. Nonetheless, for the higher threat levels (level 3-5), additional safety measures would be expected by law and subject to sanction for noncompliance. The sanction could be a monetary penalty or other strict sanction proportionate to the severity of violation [80].

All institutions and organizations would need to develop their own logistics so that they can comply with the systems effectively upon notification. For example, when a level 4 (shelter in place) alert will be dispatched, the employees would start working from home where applicable, nonessential business and schools would be closed, and people would not leave home except for the necessary activities (among other measures) until the threat level is reduced. A change in the alert level would ideally be made either in the early morning or late evening or night to prevent road congestion and panic buying. Overall, installing such a system would operate complementary to any existing platform, and a significant benefit can be conferred by enhanced public engagement.

Given that bidirectional oscillations of threat levels over the course of a pandemic may be the norm, the proposed notification system may facilitate population response changes in the direction of increasing or decreasing precautions over time as the threat accumulates or abates. In all instances, such oscillations in response recommendation should be guided first and foremost by objective scientific advice on level of health threat rather than political objectives or other nonhealth-related agendas.

Unique Challenges for Developing Countries

The administration of the graded pandemic notification system would be more challenging in developing countries for several reasons [81]. First, the accessibility of the system could be an issue due to the lack of some infrastructure components, leading to asymmetries in the reach of warnings, which may amplify the exposure and structural response disparities. Most developing countries now have well-developed mobile network systems with a majority of the population having access to cell phones, but this issue would be an ongoing consideration. As an illustration, Bangladesh, a developing country in South Asia and home of more than 164 million people [82], has reported approximately 166 million mobile phone subscribers in February 2020 [83]. Other South Asian countries have a similar level of mobile phone penetration. India is estimated to have over 800 million mobile phone users for a population of more than 1.3 billion [82,84].

Although cell phones are ubiquitous and widespread, a large proportion of people do not have reliable data plans [85]. Consequently, a notification system may need to largely rely on simple text messaging to ensure it can reach the majority of the population. Contrary to Canada where we have Alert Ready [86], mobile phone carriers in some countries do not provide the government with channels to notify the population in cases of emergency. Some countries might not have adequate technology to provide emergency alerts. Deploying such systems in developing regions, therefore, might require changes to the current infrastructure to allow for such channels to be implemented. In addition, close collaborations with cell phone providers and public health officials are necessary for the successful implementation of such technological ecosystems.

Second, the nature of each precaution level in the notification system would have to be customized to the context of developing countries [81]. For instance, the ability of large swaths of the population to enact social distancing is questionable in some countries, simply based on population density. Bangladesh and India have a population density of 1116 and 420 per square kilometer, respectively, compared to only 4 and 35 per square kilometer for Canada and the United States [82]. Therefore, a 2-meter physical distancing guideline recommended in the ongoing COVID-19 outbreak is likely not achievable.

In a similar manner, developing countries should be judicious while imposing extreme measures like lockdowns because a large proportion of the population lives under the poverty line, and the governments have limited capacity to support their citizens amid lockdown. The recent lockdown in India is worth noting as the lives of migrant workers were under threat during their efforts to return home [87]. The creation of roadway congestion could also increase the risk of disease transmission [88].

Third, the effectiveness of the notification system could be hampered by conventional wisdom or mistrust of the government. For example, a widely held belief in Mexico is that people from low socioeconomic neighborhoods benefit from greater disease immunity because they are exposed to poor sanitation standards. This belief led the governor of the Mexican State of Puebla to recently state that poor people are immune to COVID-19 [89], contrary to the fact that low standards of living are strongly associated with impaired immunity from multiple routes (eg, malnutrition, chronic stress, unhealthy behavior), leading to a greater likelihood of morbidity and mortality from COVID-19 [90-93]. The effectiveness of any notification system is prefaced on the existence of belief systems that will make the public receptive to the messages being transmitted through the system. Mistrust of government, perhaps fostered by histories of institutionalized corruption or cronyism, may also predispose the public to ignore government warnings.

In Latin America, such mistrust has led citizens to look to international sources for information on COVID-19 [94], thereby suggesting that government-sponsored apps may be disregarded altogether.

Fourth, the lessons from existing alert systems should be scrutinized. For instance, Bangladesh has been operating their Cyclone Warning System in coastal areas for decades in collaboration with the international development partners and nongovernmental organizations. This system has been successful in reducing human casualties and property damage [95,96]. However, it has been reported that a large percentage of the residents do not follow evacuation orders and do not take necessary protective measures even after receiving warning messages and being informed about the potential consequences of not evacuating [97]. The primary reasons behind this are mistrust in the warning message and failure to understand the instructions conveyed in them [97]. Nationwide implementation of a pandemic notification system, therefore, needs to address these issues carefully. The message conveyed must be simple and concise, and will need to be tailored according to the needs of the target population. To improve compliance and penetration, the simultaneous use of other means must be considered, such as announcements from churches or prayer houses, radio, television, newspaper, and other community sources.

Source and Level Considerations

An important consideration in the implementation of any pandemic notification system would be the issuing source. In this viewpoint, we have assumed that a government agency is the primary source, given the use of reserved mobile communication channels and universal population reach within a geographic region. However, it is also possible that arms-length public health organizations within or outside of governments could be considered as a notification issuing source as well. In the COVID-19 pandemic response, many central government bodies responded in a delayed manner, which would not be helped by a notification system reliant upon their approval. On the other hand, it is difficult to envision a pandemic notification system that is permitted to operate without a substantial degree of direct input from an official government source. Ultimately along these lines, a pandemic response system that is synchronized with others around the world may be ideal, particularly given the potential for rapid spread of any contagious epidemic due to air travel. Yet, on the level of any individual country or subregion within a country, substantial negotiation and multilateral approval would be required for it to operate in a manner that could reach the population with uniformity.

A final consideration is the level at which the communication is made: national, provincial or state, or regional. Given that both outbreaks and recovery rates vary significantly by subregions within a country, an argument could be made that notifications should be delivered on that level. In some countries, such geographic units have all or most of the authority to introduce such communications and make provisions to handle changes in threat level. All of these issues must be considered on a country-by-country basis. In all cases, communications must prioritize population health first and foremost.

Limitations

The main limitation of a pandemic notification system would be the unintended consequences that could occur as a result of adverse public reactions. However, this issue could be mitigated by proper education and building trust in the government. Governments, for their part, need to ensure support for all citizens to follow any recommended precautionary measures. An additional limitation would be the lack of sensitivity to local context when deployed at the national level. This is particularly important when restrictions ease if the rate of easing is different within different regions. Several other limitations have been pointed out in previous sections (eg, privacy concerns, inadequate technology, noncompliance, mistrust); however, a detailed discussion of each point is beyond the scope of this paper.

To our knowledge, this is the first paper to describe the benefits and challenges of an individual-level pandemic notification system delivered via mobile technology. We illustrated the context of why such a system is necessary and described a simple prototype model, its prospects, and potential barriers regarding implementation. However, there are other challenges that need to be addressed with the benefit of empirical data gleaned from population attitudinal surveys and real world implementational trials.

Conclusion

Following SARS-CoV and MERS-CoV, another novel coronavirus (SARS-CoV-2) emerged and rapidly spread across the world. Despite having global public health surveillance systems (eg, ProMED, GPHIN) and their timely notification on COVID-19, many nations were unsuccessful in controlling the virus effectively. To prevent or slow the progression of future pandemics, we need a remolded system capable of facilitating effective communication and prompt activation of preventive measures. A graded, individual-level pandemic notification system could be an effective tool if customized and carefully implemented on the level of the individual country. In this system, the threat levels would be arranged hierarchically with relevant logos and a set of recommended precautionary measures for each level, which would be easily comprehensible to all citizens. Notifications provided by such a system could facilitate a prompt, timely, and coordinated response from all levels of governments and, most importantly, from the public. However, extensive preparation, advance education, and coordination by the government with all the relevant sectors involved in the dissemination of the alert are required to prevent unintended consequences. Public education and awareness are of utmost importance to ensure that the alert is understood and appreciated properly upon delivery. Under such circumstances, citizens would be able to act promptly in an organized manner relevant to the alert level without having unproductive panic and fear. It is worth noting that different countries and regions have their own limitations and challenges. Implementing a pandemic notification system would not be effective unless authorities address unique challenges in the local context. Particularly for the developing regions, careful consideration of limitations and potential unintended consequences is required.

Acknowledgments

Support for this research was provided by operating grants to PH (RGPIN-2019-04606) and PM (RGPIN-2017-05310) from the Natural Sciences and Engineering Council of Canada, as well as a foundation grant (FDN-148477) from the Canadian Institutes of Health Research, and a Senior Investigator Award (IA-004) from the Ontario Institute for Cancer Research to GTF.

Authors' Contributions

All authors contributed significantly to the conceptualization of this paper, provided critical feedback, and contributed to the drafting and revision of the article. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
2. Gao G. From "A"IV to "Z"IKV: attacks from emerging and re-emerging pathogens. Cell 2018 Mar 08;172(6):1157-1159. [doi: 10.1016/j.cell.2018.02.025] [Medline: 29522735]
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
4. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. China CDC Weekly 2020;2(8):113-122.
5. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
6. Johns Hopkins University. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) URL: https://coronavirus.jhu.edu/map.html [accessed 2020-05-31]
7. World Health Organization. Rolling updates on coronavirus disease (COVID-19) URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2020-05-06]
8. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
9. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020 Feb 07:1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
10. World Health Organization. 2020 Mar 13. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected URL: https://www.who.int/docs/default-source/coronaviruse/clinical-management-of-novel-cov.pdf [accessed 2020-05-06]
11. Lipsitch M. Estimating case fatality rates of COVID-19. Lancet Infect Dis 2020 Mar:1. [doi: 10.1016/s1473-3099(20)30245-0]
12. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis 2020 Jun;20(6):669-677. [doi: 10.1016/S1473-3099(20)30243-7]
13. Khafaie MA, Rahim F. Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. Osong Public Health Res Perspect 2020 Apr;11(2):74-80 [FREE Full text] [doi: 10.24171/j.phrp.2020.11.2.03] [Medline: 32257772]
14. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA 2020 Mar 23:e. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]
15. World Health Organization. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003 URL: https://www.who.int/csr/sars/country/table2004_04_21/en/ [accessed 2020-05-06]
16. World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV) URL: https://www.who.int/emergencies/mers-cov/en/ [accessed 2020-05-06]
17. Wilder-Smith A, Chiew CJ, Lee VJ. Can we contain the COVID-19 outbreak with the same measures as for SARS? Lancet Infect Dis 2020 May;20(5):e102-e107. [doi: 10.1016/s1473-3099(20)30129-8]
18. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020 Mar 13;27(2):1-4 [FREE Full text] [doi: 10.1093/jtm/taaa021] [Medline: 32052846]
19. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]

20. Mogi R, Spijker J. The influence of social and economic ties to the spread of COVID-19 in Europe. SocArXiv 2020:1-27. [doi: 10.31235/osf.io/sb8xn]
21. Rodriguez-Morales A, Gallego V, Escalera-Antezana J, Méndez CA, Zambrano L, Franco-Paredes C, et al. COVID-19 in Latin America: the implications of the first confirmed case in Brazil. Travel Med Infect Dis 2020 Feb 29:101613 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101613] [Medline: 32126292]
22. Saker L, Lee K, Cannito B, Gilmore A, Campbell-Lendrum D. Globalization and infectious diseases: a review of the linkages. World Health Organization 2004:a.
23. Carrion M, Madoff L. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. Int Health 2017 May 01;9(3):177-183 [FREE Full text] [doi: 10.1093/inthealth/ihx014] [Medline: 28582558]
24. ProMED: International Society for Infectious Diseases. 2020. URL: https://promedmail.org/ [accessed 2020-05-06]
25. Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and Early Warning Outbreak Detection. Can J Public Health 2006 Jan 1;97(1):42-44. [doi: 10.1007/bf03405213]
26. Factiva. URL: https://professional.dowjones.com/factiva/ [accessed 2020-05-06]
27. Al Bawaba. 2020. URL: https://www.albawaba.com/ [accessed 2020-05-06]
28. Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clin Microbiol Infect 2014 Jan;20(1):17-21. [doi: 10.1111/1469-0691.12477] [Medline: 24350723]
29. Mackenzie JS, Drury P, Arthur RR, Ryan MJ, Grein T, Slattery R, et al. The global outbreak alert and response network. Glob Public Health 2014;9(9):1023-1039. [doi: 10.1080/17441692.2014.951870] [Medline: 25186571]
30. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
31. Brewster M. CBC. 2020 Apr 22. Inside Canada's frayed pandemic early warning system and its COVID-19 response URL: https://www.cbc.ca/news/politics/covid19-pandemic-early-warning-1.5537925 [accessed 2020-05-06]
32. Chang S, Harding N, Zachreson C, Cliff O, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. arXiv preprint 2020:e.
33. Goniewicz K, Khorram-Manesh A, Hertelendy A, Goniewicz M, Naylor K, Burkle F. Current response and management decisions of the European Union to the COVID-19 outbreak: a review. Sustainability 2020 May 08;12(9):3838. [doi: 10.3390/su12093838]
34. Schabas R. SARS: prudence, not panic. CMAJ 2003 May 27;168(11):1432-1434 [FREE Full text] [Medline: 12771074]
35. Government of Canada. 2014 Sep 04. Special theme issue: Ebola preparedness in Canada URL: https://www.canada.ca/en/public-health/services/reports-publications/canada-communicable-disease-report-ccdr/monthly-issue/2014-40/ccdr-volume-40-15-september-4-2014.html [accessed 2020-05-16]
36. Newman M. Covid-19: doctors' leaders warn that staff could quit and may die over lack of protective equipment. BMJ 2020 Mar 26;368:m1257. [doi: 10.1136/bmj.m1257] [Medline: 32217522]
37. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/NEJMsb2005114] [Medline: 32202722]
38. Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. BMJ 2020 Mar 18;368:m1090. [doi: 10.1136/bmj.m1090] [Medline: 32188598]
39. Armocida B, Formenti B, Ussai S, Palestra F, Missoni E. The Italian health system and the COVID-19 challenge. Lancet Public Health 2020 May;5(5):e253. [doi: 10.1016/s2468-2667(20)30074-8]
40. Verelst F, Kuylen E, Beutels P. Indications for healthcare surge capacity in European countries facing an exponential increase in coronavirus disease (COVID-19) cases, March 2020. Eurosurveillance 2020 Apr 02;25(13):e. [doi: 10.2807/1560-7917.es.2020.25.13.2000323]
41. Government of Canada. Coronavirus disease (COVID-19) URL: https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html [accessed 2020-05-06]
42. Vogel L. Viral misinformation threatens public health. CMAJ 2017 Dec 18;189(50):E1567 [FREE Full text] [doi: 10.1503/cmaj.109-5536] [Medline: 29255104]
43. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
44. Farooq A, Laato S, Islam AKMN. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]
45. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
46. Cuan-Baltazar J, Muñoz-Perez M, Robledo-Vega C, Pérez-Zepeda M, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Preprints 2020 Feb 26;6(2):e18444. [doi: 10.2196/preprints.18444]

47. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
48. Budhwani H, Sun R. Creating COVID-19 stigma by referencing the novel coronavirus as the "Chinese virus" on Twitter: quantitative analysis of social media data. J Med Internet Res 2020 May 06;22(5):e19301 [FREE Full text] [doi: 10.2196/19301] [Medline: 32343669]
49. Martins-Filho PR, Santos VS. No evidence supports the use of ether and chloroform inhalation for treating COVID-19. Rev Panam Salud Publica 2020;44:e41 [FREE Full text] [doi: 10.26633/RPSP.2020.41] [Medline: 32269593]
50. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
51. Trew B. Independent. 2020 Mar 27. Coronavirus: hundreds dead in Iran from drinking methanol amid fake reports it cures disease URL: https://www.independent.co.uk/news/world/middle-east/iran-coronavirus-methanol-drink-cure-deaths-fake-a9429956.html [accessed 2020-05-06]
52. Stickings T. Daily Mail. 2020 Mar 19. Coronavirus patients line the corridors in footage from inside Italian hospital, as military trucks transport scores of victims' coffins to be cremated URL: https://www.dailymail.co.uk/news/article-8129959/Military-trucks-transport-Italian-coronavirus-victims-coffins-cremated.html [accessed 2020-05-06]
53. Gallón N. CNN. 2020 Apr 03. Bodies are being left in the streets in an overwhelmed Ecuadorian city URL: https://www.cnn.com/2020/04/03/americas/guayaquil-ecuador-overwhelmed-coronavirus-intl/index.html [accessed 2020-05-06]
54. Leung CC, Lam TH, Cheng KK. Mass masking in the COVID-19 epidemic: people need guidance. Lancet 2020 Mar;395(10228):945. [doi: 10.1016/s0140-6736(20)30520-1]
55. La VP, Pham TH, Ho MT, Hoàng N, Linh N, Trang VT, et al. Policy response, social media and science journalism for the sustainability of the public health system amid COVID-19 outbreak: the Vietnam lessons. OSF Preprints 2020 Mar 22;12(7):2931. [doi: 10.31219/osf.io/vxhz5]
56. Sagan SD. Nuclear alerts and crisis management. Int Secur 1985;9(4):99. [doi: 10.2307/2538543]
57. Tsang NKF, Wong O, Prideaux B. An evaluation of the effectiveness of travel advisories with a specific focus on Hong Kong's outbound travel alert system. J Vacation Marketing 2018 Oct 05;24(4):307-323. [doi: 10.1177/1356766717725563]
58. Baker S, Bloom N, Davis S, Terry S. Covid-induced economic uncertainty. Natl Bureau of Econ Res 2020:1-16. [doi: 10.3386/w26983]
59. Baldwin R, di Mauro BW. Economics in the Time of COVID-19. New York: Centre for Economic Policy Research; Mar 06, 2020.
60. Draaisma M. CBC. 2020 Jan 12. Ontario government apologizes for alert about Pickering nuclear plant incident sent 'in error' URL: https://www.cbc.ca/news/canada/toronto/pickering-nuclear-generating-station-1.5424115 [accessed 2020-05-06]
61. Kepner RM. Efficiency of the emergency alert system. Washington State University 2010 Dec:1-166.
62. Jin D, Lin J. Managing tsunamis through early warning systems: a multidisciplinary approach. Ocean Coastal Manage 2011 Feb;54(2):189-199. [doi: 10.1016/j.ocecoaman.2010.10.025]
63. Bernard E, Titov V. Evolution of tsunami warning systems and products. Philos Trans A Math Phys Eng Sci 2015 Oct 28;373(2053):e [FREE Full text] [doi: 10.1098/rsta.2014.0371] [Medline: 26392620]
64. Cooke A. CBC. 2020 Apr 21. 'I would not have let my wife leave': widower blasts lack of emergency alert during N.S. shootings URL: https://www.cbc.ca/news/canada/nova-scotia/nova-scotia-shooting-emergency-alert-1.5540220 [accessed 2020-05-06]
65. Brown C. CBC. 2020 Apr 22. New Zealand's COVID-19 'eradication' strategy looks to be paying off URL: https://www.cbc.ca/news/world/covid-new-zealand-eradicate-chris-brown-cbc-1.5540021 [accessed 2020-05-06]
66. Hollingsworth J. CNN. 2020 Apr 09. New Zealand reported a decline in new coronavirus cases for the fourth consecutive day. The country is still tightening its border restrictions URL: https://www.cnn.com/2020/04/09/asia/new-zealand-lessons-intl-hnk/index.html [accessed 2020-05-06]
67. New Zealand Government. COVID-19 alert system URL: https://covid19.govt.nz/alert-system/covid-19-alert-system/ [accessed 2020-05-06]
68. Worldometer. 2020. COVID-19 Coronavirus Pandemic URL: https://www.worldometers.info/coronavirus/ [accessed 2020-05-31]
69. Sabbagh D. The Guardian. 2020 May 10. Boris Johnson announces five-tier coronavirus alert system URL: https://www.theguardian.com/politics/2020/may/10/boris-johnson-to-announce-five-tier-coronavirus-alert-system [accessed 2020-05-25]
70. The Telegraph. 2020 May 25. What are the five levels of coronavirus alert in the UK? URL: https://www.telegraph.co.uk/news/0/five-levels-covid-19-coronavirus-uk-alert/ [accessed 2020-05-25]
71. Tsekouras P. CTV News. 2020 Apr 05. Ontario issues another emergency alert on COVID-19, province confirms 27 new deaths URL: https://toronto.ctvnews.ca/ontario-issues-another-emergency-alert-on-covid-19-province-confirms-27-new-deaths-1.4882526 [accessed 2020-05-06]
72. The Canadian Press. 2020 Apr 20. Alert Bay's tsunami siren signals curfew to fight COVID-19, says B.C. mayor URL: https://www.alaskahighwaynews.ca/alert-bay-s-tsunami-siren-signals-curfew-to-fight-covid-19-says-b-c-mayor-1.24121215 [accessed 2020-05-06]

73. Soupcoff M. National Post. 2020 Mar 30. We need better, clearer communication from our governments on COVID-19 URL: https://nationalpost.com/opinion/marni-soupcoff-we-need-better-clearer-communication-from-our-governments-on-covid-19 [accessed 2020-05-06]
74. Tumilty R. National Post. 2020 Mar 23. COVID-19: Trudeau to Canadians: 'enough is enough. Go home and stay home' URL: https://nationalpost.com/news/canada/covid-19-ontario-reports-78-new-cases-the-most-in-one-day-so-far [accessed 2020-05-06]
75. Malik A. MobileSyrup. 2020 Apr 06. About 84 percent of Canadians want to receive COVID-19 updates via alert system: survey URL: https://mobilesyrup.com/2020/04/06/national-alerts-covid-19-survey/ [accessed 2020-05-06]
76. Kelso JK, Milne GJ, Kelly H. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. BMC Public Health 2009 Apr 29;9:117 [FREE Full text] [doi: 10.1186/1471-2458-9-117] [Medline: 19400970]
77. Bootsma MCJ, Ferguson NM. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. Proc Natl Acad Sci U S A 2007 May 01;104(18):7588-7593 [FREE Full text] [doi: 10.1073/pnas.0611071104] [Medline: 17416677]
78. Andersen M. Early evidence on social distancing in response to COVID-19 in the United States. SSRN Journal 2020:e. [doi: 10.2139/ssrn.3569368]
79. Karako K, Song P, Chen Y, Tang W. Analysis of COVID-19 infection spread in Japan based on stochastic transition model. Biosci Trends 2020 May 21;14(2):134-138. [doi: 10.5582/bst.2020.01482] [Medline: 32188819]
80. McCreadie D. CBC. 2020 Apr 05. Education over enforcement: what to expect under new physical distancing rules URL: https://www.cbc.ca/news/canada/new-brunswick/physical-distancing-enforcement-rcmp-1.5522736 [accessed 2020-05-06]
81. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle P, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. Lancet 2020 Mar;395(10227):871-877. [doi: 10.1016/s0140-6736(20)30411-6]
82. World Population Review. 2020 world population by country URL: https://worldpopulationreview.com/ [accessed 2020-05-06]
83. Bangladesh Telecommunication Regulatory Commission. Mobile phone subscribers in Bangladesh February, 2020 URL: http://www.btrc.gov.bd/content/mobile-phone-subscribers-bangladesh-february-2020 [accessed 2020-05-06]
84. Statista. Number of mobile phone users in India from 2013 to 2019 URL: https://www.statista.com/statistics/274658/forecast-of-mobile-phone-users-in-india/ [accessed 2020-05-06]
85. Castelo M. Half the world without Internet: how brands can boost access to data on mobile phones. J Brand Strategy 2016;5(3):294-297.
86. Alert Ready. URL: https://www.alertready.ca/ [accessed 2020-05-06]
87. Biswas S. BBC. 2020 Mar 30. Coronavirus: India's pandemic lockdown turns into a human tragedy URL: https://www.bbc.com/news/world-asia-india-52086274 [accessed 2020-05-06]
88. Saha N. News Ghana. 2020 Apr 16. Bangladesh is about to lose its battle against Covid-19 pandemic URL: https://newsghana.com.gh/bangladesh-is-about-to-lose-its-battle-against-covid-19-pandemic/ [accessed 2020-05-06]
89. Chakraborty B. New York Post. 2020 Mar 27. Mexican governor sparks outrage after claiming poor are ‘immune’ to coronavirus URL: https://nypost.com/2020/03/27/mexican-governor-sparks-outrage-after-claiming-poor-are-immune-to-coronavirus/ [accessed 2020-05-06]
90. Nandi A, Glymour MM, Subramanian SV. Association among socioeconomic status, health behaviors, and all-cause mortality in the United States. Epidemiology 2014;25(2):170-177. [doi: 10.1097/ede.0000000000000038]
91. Bourke C, Berkley J, Prendergast A. Immune dysfunction as a cause and consequence of malnutrition. Trends Immunol 2016 Jun;37(6):386-398. [doi: 10.1016/j.it.2016.04.003] [Medline: 27237815]
92. Ziol-Guest KM, Duncan GJ, Kalil A, Boyce WT. Early childhood poverty, immune-mediated disease processes, and adult productivity. Proc Natl Acad Sci U S A 2012 Oct 16;109 Suppl 2:17289-17293 [FREE Full text] [doi: 10.1073/pnas.1203167109] [Medline: 23045664]
93. Abrams EM, Szefler SJ. COVID-19 and the impact of social determinants of health. Lancet Respir Med 2020 May:1-2. [doi: 10.1016/s2213-2600(20)30234-4]
94. Rubio D. Premise. Latin American countries lack trust in government information reveals premise URL: https://www.premise.com/latin-american-countries-lack-trust-in-government-information-reveals-premise/ [accessed 2020-05-06]
95. Akhand MH. Disaster management and cyclone warning system in Bangladesh. In: Zschau J, Küppers A, editors. Early Warning Systems for Natural Disaster Reduction. Berlin, Heidelberg: Springer; 2003:49-64.
96. Haque U, Hashizume M, Kolivras KN, Overgaard HJ, Das B, Yamamoto T. Reduced death rates from cyclones in Bangladesh: what more needs to be done? Bull World Health Organ 2011 Oct 24;90(2):150-156. [doi: 10.2471/blt.11.088302]
97. Roy C, Sarkar SK, Åberg J, Kovordanyi R. The current cyclone early warning system in Bangladesh: providers' and receivers' views. Int J Disaster Risk Reduction 2015 Jun;12:285-299. [doi: 10.1016/j.ijdrr.2015.02.004]

Abbreviations

COVID-19: coronavirus disease
GPHIN: Global Public Health Intelligence Network
MERS: Middle East respiratory syndrome
MERS-CoV: Middle East respiratory syndrome–related coronavirus
ProMED: Program for Monitoring Emerging Diseases
R_0: basic reproduction number
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome–related coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by S Khan, F Holl, K Goniewicz; comments to author 25.05.20; revised version received 28.05.20; accepted 29.05.20; published 05.06.20.

Please cite as:
Sakib MN, Butt ZA, Morita PP, Oremus M, Fong GT, Hall PA
Considerations for an Individual-Level Population Notification System for Pandemic Response: A Review and Prototype
J Med Internet Res 2020;22(6):e19930
URL: http://www.jmir.org/2020/6/e19930/
doi:10.2196/19930
PMID:32484443

Original Paper

COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan

Peng-Wei Wang[1,2*], MD, PhD; Wei-Hsin Lu[3,4*], MD; Nai-Ying Ko[5], RN, PhD; Yi-Lung Chen[6,7], PhD; Dian-Jeng Li[8,9], MD; Yu-Ping Chang[10], RN, PhD; Cheng-Fang Yen[1,2*], MD, PhD

[1]Department of Psychiatry, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan
[2]Department of Psychiatry, School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[3]Department of Psychiatry, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chia-Yi, Taiwan
[4]Department of Senior Citizen Service Management, Chia Nan University of Pharmacy and Science, Tainan, Taiwan
[5]Department of Nursing, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[6]Department of Healthcare Administration, Asia University, Taichung, Taiwan
[7]Department of Psychology, Asia University, Taichung, Taiwan
[8]Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[9]Department of Addiction Science, Kaohsiung Municipal Kai-Syuan Psychiatric Hospital, Kaohsiung, Taiwan
[10]School of Nursing, The State University of New York, University at Buffalo, Buffalo, NY, United States
[*]these authors contributed equally

Corresponding Author:
Cheng-Fang Yen, MD, PhD
Department of Psychiatry
Kaohsiung Medical University Hospital
Kaohsiung Medical University
100 Tzyou 1st Road
Kaohsiung, 807
Taiwan
Phone: 886 7 3121101 ext 6822
Email: chfaye@cc.kmu.edu.tw

Abstract

Background: People obtain information on the coronavirus disease (COVID-19) from the internet and other sources. Understanding the factors related to such information sources aids health professionals in educating individuals.

Objective: This study used data collected from the online survey study on COVID-19 in Taiwan to examine what major COVID-19 information sources are available and which sources are significantly related to the self-confidence of people in coping with COVID-19 in Taiwan.

Methods: A total of 1904 participants (1270 non–health-care workers and 634 health care workers) were recruited from the Facebook advertisement. Their major sources of information about COVID-19, the relationships between the sources and demographic factors, and the relationships between the sources and the self-confidence in coping with COVID-19 were surveyed.

Results: Most Taiwanese people relied on the internet for COVID-19 information. Many respondents also used a variety of sources of information on COVID-19; such variety was associated with sex, age, and the level of worry toward COVID-19, as well as if one was a health care worker. For health care workers, the use of formal lessons as an information source was significantly associated with better self-confidence in coping with COVID-19. The significant association between receiving information from more sources and greater self-confidence was found only in health care workers but not in non–health-care workers.

Conclusions: Medical professionals should consider subgroups of the population when establishing various means to deliver information on COVID-19.

(J Med Internet Res 2020;22(6):e20021) doi:10.2196/20021

KEYWORDS

COVID-19; information; internet; coping; confidence; mental health; social media; Facebook; survey; online health information

Introduction

The coronavirus disease (COVID-19) emerged in Wuhan, China at the end of 2019 and rapidly spread worldwide. The World Health Organization (WHO) declared the ongoing outbreak of COVID-19 as a public health emergency demanding global concern. Early in the outbreak, Taiwan was predicted to have the second highest number of COVID-19 cases due to the many people moving to and from Taiwan and China [1]. After all, Taiwan experienced the severe outbreak of the 2002-2003 severe acute respiratory syndrome (SARS), which originated from southern China; worldwide, Taiwan had the third highest number of SARS cases, after China and Hong Kong [2]. However, the experience with SARS made many Taiwanese people vigilant against COVID-19, which has aided in COVID-19 prevention.

For information on COVID-19, timeliness and accuracy—although difficult to attain and measure—are foundational to mitigating and curing the disease for both the public and the scientific community [3]. Among the public, the internet is the most popular source of information on the etiologies and intervention models of medical illnesses [4]. Information has been proliferating on traditional and social media since the COVID-19 outbreak [5]. A recent study of 21 countries found that the number of Google searches for “wash hands” increased with the lower speed of the COVID-19 spread [6]. Misinformation on COVID-19, however, has also been proliferating on the internet, especially on social media [7,8].

In addition to the internet, traditional media are also important sources of information during disease outbreaks. However, repeated media exposure to crisis-related information elevates anxiety and stress responses among people [9]. The public may also receive information on COVID-19 from medical staff and laypeople, such as their friends, family members, and coworkers. Because people obtain information on COVID-19 from various sources, understanding the factors related to such information sources aids health professionals in educating individuals in particular and the public at large. Thus, by developing information-delivery systems that are transparent and effective, self-confidence in coping with the pandemic can be improved in their audience.

The online survey study on COVID-19 in Taiwan (hereafter referred to as the Taiwan COVID-19 survey) was conducted to assess the life experiences of people in Taiwan during the COVID-19 outbreak. Generally, the online survey is a promising method for assessing how members of the public understand and perceive a fast-moving infectious disease outbreak [10]. This study used data collected from the Taiwan COVID-19 survey to examine the following two issues: (1) what major COVID-19 information sources are available as well as the relationships between these sources and demographic factors; and (2) which sources are significantly related to the self-confidence of people in coping with COVID-19? We were particularly interested in the information sources of traditional media and the internet with regard to advice on social distancing. Furthermore, because Taiwan’s Ministry of Health and Welfare recommends that health care workers should learn about COVID-19 through online or in-person lessons, we were also interested in whether the associations of information sources with people’s self-confidence differed between health care workers and non–health-care workers.

Methods

Participants

Participants were recruited through a Facebook advertisement from April 10 to April 20, 2020. Facebook users were eligible for this study if they were ≥20 years of age and living in Taiwan. The Facebook advertisement included a headline, main text, pop-up banner, and link to the research questionnaire website. We designed the advertisement to appear in the Facebook users’ “news feeds,” which is a continually updated list from advertisers and the user’s connections such as friends and the Facebook groups that they have joined. Our advertisement only targeted users’ news feeds as opposed to other Facebook advertising locations such as on the right column because news feed advertisements are most effective in recruiting research participants [11]. We targeted the advertisement to Facebook users by location (Taiwan) and language (Chinese), where Facebook’s advertising algorithm determined which users to show our advertisement to. To ensure that health care workers were recruited, we also posted the link of the Facebook advertisement to Line (a direct messaging app) and Facebook groups joined by health care workers.

This study was approved by the Institutional Review Board (IRB) of Kaohsiung Medical University Hospital (KMUHIRB-EXEMPT[I] 20200011). Because participation was voluntary and survey responses were anonymous, the IRB ruled that this study did not require informed consent. Our study participants were given no incentive for participation. We provided links to COVID-19 information from the Taiwan Centers for Disease Control, Kaohsiung Medical University Hospital, and Medical College of National Cheng Kung University for participants to learn more about COVID-19.

Measures

In the survey, participants were asked about the following:

Sources of Information About COVID-19

We measured participants’ frequency of receiving COVID-19-related information from the following sources: the internet (including blogs, internet news, and social media, such as Facebook, Line, Twitter, and Plurk), friends, traditional media (including television, newspapers, and radio broadcasting), formal lessons on COVID-19 (whether online or in-person), medical staff in health care settings, coworkers, and family members. Participants rated their frequency for each source as never, sometimes, or frequently. Sources rated frequently were classified as major sources of information. The total numbers of sources of information about COVID-19 were summed.

Worry About COVID-19

We used the following question from the study of Liao et al [12] to assess how much respondents worried about COVID-19: “Please rate the level of your current worry toward COVID-19.” This question was rated from 1 (very mild) to 10 (very severe).

Self-Confidence in Coping With COVID-19

We used the following question adapted from the questionnaire on risk perception of an infectious disease outbreak [13] to assess how self-confident respondents were in coping with COVID-19: “How confident are you that you can cope well with COVID-19?” The question was rated from 1 (not confident at all) to 5 (very confident).

Demographic Variables

Data on respondents’ gender (female and male), age, and education level were also collected. Respondents who had high school qualifications or below and respondents who had university qualifications or above were classified into low and high education level groups, respectively. Respondents were also asked whether they were health care workers.

Statistical Analysis

Data analysis was performed using SPSS 22.0 statistical software (IBM Corp). For each participant, the frequency of information-source use is expressed in terms of percentage. The associations of demographic characteristics and worry toward COVID-19 with COVID-19 information sources were examined using logistic regression analysis. The *P* value, odds ratio, and 95% CI were used to indicate significance. Because multiple comparisons were conducted, a *P* value of <.007 (.05/7) indicated significance. The factors related to the total number of major sources of information about COVID-19 were examined using multiple regression analysis. Moreover, the associations of each information source and total number of information sources with self-confidence were examined using multiple regression analyses as well as for the subgroups of health care workers and non–health-care workers. A two-tailed *P* value of <.05 indicated statistical significance.

Results

In total, the data of 1904 respondents (1282 female and 622 male participants) were analyzed, with 70 of the original 1974 respondents excluded due to missing data. Table 1 presents the descriptive statistics for demographic characteristics, worry toward COVID-19, self-confidence in coping with COVID-19, and COVID-19 information sources. The mean age was 38.0 years (SD 10.8 years); most had a high education level; one-third were health care workers. The mean scores for worry and self-confidence were 6.1 (SD 2.2) and 3.1 (SD 0.8), respectively.

Table 1. Descriptive statistics for participants' (N=1904) demographic characteristics, worry toward COVID-19, self-confidence in coping with COVID-19, and COVID-19 information sources.

Variables	Statistic	Range
Gender, n (%)		
Female	1282 (67.3)	N/A[a]
Male	622 (32.7)	N/A
Age (years), mean (SD)	38.2 (10.8)	20-74
Education level, n (%)		
Low (high school or below)	212 (11.1)	N/A
High (university or above)	1692 (88.9)	N/A
Health care workers, n (%)		
No	1270 (66.7)	N/A
Yes	634 (33.3)	N/A
Severity of worry toward COVID-19[b], mean (SD)	6.1 (2.2)	1-10
Self-confidence in coping with COVID-19, mean (SD)	3.4 (0.8)	1-5
Sources of information about COVID-19, n (%)		
Internet	1535 (80.6)	N/A
Friends	403 (21.2)	N/A
Traditional media	1018 (53.5)	N/A
Academic lessons	391 (20.5)	N/A
Medical staff	367 (19.3)	N/A
Work colleagues	462 (24.3)	N/A
Family members	475 (24.9)	N/A
Number of information sources, n (%)		
0	200 (10.5)	N/A
1	494 (25.9)	N/A
2	453 (23.8)	N/A
3	284 (14.9)	N/A
4	196 (10.3)	N/A
5	129 (6.8)	N/A
6	66 (3.5)	N/A
7	82 (4.3)	N/A
Total number of information sources, mean (SD)	2.4 (1.8)	0-7

[a] Not applicable.

[b] COVID-19: coronavirus disease.

The major source of information on COVID-19 in a high proportion of users was the internet, followed by traditional media, family members, coworkers, friends, formal lessons, and medical staff. Furthermore, more than one-quarter of participants received their COVID-19-related information from 1 source, followed by 2 sources, 3 sources, ≥5 sources, 0 sources, and 4 sources, with 2.4 sources as the mean (SD 1.8).

Table 2 presents the multiple regression results on the associations of demographic characteristics and worry toward COVID-19 with each information source. The results indicated that women were more likely to obtain information from traditional media and family members; older people were more likely to obtain information from traditional media and were less likely to obtain information from the internet and their work colleagues; and health care workers were more likely to obtain information from formal lessons, medical staff, and coworkers. For all information sources, except for formal lessons and medical staff, worry was significantly associated with information-source use.

Table 2. Factors related to COVID-19 information sources.

Variables	Internet		Friends		Traditional media		Academic lessons		Medical staff		Work colleagues		Family members	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Males[b]	0.814 (0.638-1.040)	.10	1.008 (0.790-1.286)	.95	0.724 (0.594-0.883)	.001	0.939 (0.724-1.218)	.64	0.818 (0.618-1.084)	.16	0.741 (0.579-0.948)	.02	0.718 (0.566-0.911)	.006
Age	0.977 (0.967-0.988)	<.001	0.987 (0.976-0.998)	.02	1.016 (1.007-1.025)	<.001	1.000 (0.988-1.012)	>.99	0.991 (0.979-1.004)	.18	0.983 (0.972-0.994)	.002	0.997 (0.986-1.007)	.53
High educational level[c]	1.545 (1.097-2.176)	.01	1.016 (0.701-1.474)	.93	0.822 (0.608-1.111)	.20	0.995 (0.653-1.515)	.98	0.683 (0.441-1.058)	.09	0.706 (0.495-1.008)	.06	0.712 (0.515-0.983)	.04
Health care workers[d]	0.827 (0.647-1.057)	.13	1.309 (1.032-1.661)	.03	0.992 (0.814-1.210)	.94	5.891 (4.600-7.543)	<.001	10.645 (8.067-14.046)	<.001	3.500 (2.783-4.402)	<.001	0.827 (0.655-1.045)	.11
Worry toward COVID-19[e]	1.107 (1.052-1.165)	<.001	1.107 (1.051-1.165)	<.001	1.094 (1.049-1.140)	<.001	1.013 (0.959-1.069)	.64	1.065 (1.004-1.129)	.04	1.084 (1.031-1.141)	.002	1.102 (1.050-1.156)	<.001

[a] OR: odds ratio.

[b] Female as reference.

[c] High school or below as reference.

[d] Non–health-care worker as reference.

[e] COVID-19: coronavirus disease.

Table 3 presents the multiple regression results on the associations of demographic characteristics and worry toward COVID-19 with the number of information sources each respondent used. The results indicated that women and health care workers tended to use more information sources relative to their counterparts. Worry was also significantly associated with the number of information sources used.

Table 3. Factors related to number of COVID-19 information sources.

Variables	Beta	*t* test	*P* value
Males[a]	–.066	–2.909	.004
Age	–.037	–1.585	.11
High educational level[b]	–.024	–1.067	.29
Health care workers[c]	.117	5.218	<.001
Worry toward COVID-19[d]	.230	10.132	<.001

[a] Female as reference.

[b] High school or below as reference.

[c] Non–health-care worker as reference.

[d] COVID-19: coronavirus disease.

Table 4 presents the multiple regression results on the associations of each information source and the number of information sources used by a respondent with self-confidence for the health care worker and non–health-care worker subgroups. For health care workers, those who received information from formal lessons and from more sources had significantly greater self-confidence. For non–health-care workers, a nonsignificant association was noted between each information source and the total number of sources used by a participant with self-confidence.

Table 4. Associations of COVID-19 information sources with self-confidence in coping with COVID-19.

Variables	Non–health-care workers						Health care workers					
	Beta	*t* test	*P* value	Beta	*t* test	*P* value	Beta	*t* test	*P* value	Beta	*t* test	*P* value
Males[a]	.082	3.030	.002	.081	3.009	.003	.069	1.781	.08	.089	2.279	.02
Age	.002	0.088	.93	.002	0.054	.96	–.008	–0.205	.84	–.015	–0.401	.69
High educational level[b]	.052	1.902	.06	.050	1.854	.06	–.001	–0.039	.97	.014	0.361	.72
Worry toward COVID-19[c]	–.307	–11.258	<.001	–.309	–11.363	<.001	–.323	–8.525	<.001	–.333	–8.747	<.001
Information from the internet	.001	0.051	.96	N/A[d]	N/A	N/A	.064	1.552	.12	N/A	N/A	N/A
Information from friends	–.027	–0.906	.37	N/A	N/A	N/A	.019	0.435	.66	N/A	N/A	N/A
Information from traditional media	–.022	–0.767	.44	N/A	N/A	N/A	–.072	–1.628	.10	N/A	N/A	N/A
Information from academic lessons	.049	1.659	.097	N/A	N/A	N/A	.144	2.969	.003	N/A	N/A	N/A
Information from medical staff	.024	0.782	.44	N/A	N/A	N/A	.085	1.608	.11	N/A	N/A	N/A
Information from work colleagues	–.016	–0.492	.62	N/A	N/A	N/A	–.081	–1.644	.10	N/A	N/A	N/A
Information from family members	.052	1.703	.09	N/A	N/A	N/A	–.042	–0.986	.32	N/A	N/A	N/A
Total number of information sources	N/A	N/A	N/A	.035	1.296	.20	N/A	N/A	N/A	.078	2.041	.04

[a] Female as reference.

[b] High school or below as reference.

[c] COVID-19: coronavirus disease.

[d] Not applicable.

Discussion

Principal Findings

This online survey study found that most Taiwanese people relied on the internet for COVID-19 information, followed by traditional media. Many respondents also used a variety of sources of information on COVID-19; such variety was associated with sex, age, and the level of worry toward COVID-19, as well as if one was a health care worker. For health care workers, the use of formal lessons as an information source was significantly associated with better self-confidence in coping with COVID-19. The significant association between receiving information from more sources and greater self-confidence was found only in health care workers but not in non–health-care workers.

The Internet and Traditional Media as COVID-19 Information Sources

Approximately 80% of participants received COVID-19 information online. The internet makes information on COVID-19 more accessible, especially for those staying indoors due to the pandemic, with the websites of official public health organizations being the highest-quality source of online information on COVID-19 and how to prevent it [14]. We found that age was positively and negatively related with receiving COVID-19 information from traditional media and from the internet, respectively. This result suggests a cohort effect on the information source related to major health issues. Because the internet is a popular and accessible information source, medical professionals should consider subgroups of the population, especially with respect to internet access, when delivering information online. Despite the popularity and accessibility of the internet, we did not find a significant association between use of the internet as an information source on COVID-19 and self-confidence in coping with COVID-19. Misinformation on COVID-19 is rife, especially on social media. The WHO and Taiwan's Ministry of Health and Welfare made efforts to dispel such misinformation [7,8].

Sources of Information About COVID-19 Among Health Care Workers

Because Taiwan's Ministry of Health and Welfare recommended health care workers attend formal lessons (whether online or in-person) on COVID-19, they were more likely to receive COVID-19 information from formal lessons, which resulted in greater self-confidence. A study in China reported depression, anxiety, insomnia, and distress in a high proportion of health care workers who were never exposed to information on COVID-19 [15]. It was the uncertainty surrounding COVID-19, in addition to physical and psychological exhaustion, that resulted in such mental health problems [16]. Thus, our study and previous studies have demonstrated the necessity of providing timely and transparent formal lessons on COVID-19 for health care workers.

Worry About COVID-19 and Sources of Information

We noted that more severe worry about COVID-19 was significantly associated with using more information sources. More sources of information about COVID-19 were also associated with higher self-confidence to cope with COVID-19 in health care workers. Overall, these results imply the benefits of using multiple information sources during the COVID-19 pandemic. However, the direction of the association is unclear: greater self-confidence can lead one to access information from multiple sources. Clarifying this direction of association warrants further study.

Limitations

This study has some limitations. First, although recruiting participants through Facebook is a promising research method to target the public during fast-moving infectious disease outbreaks [10], Facebook users may not be representative of the population. A review of a study that recruited participants through Facebook reported a bias in favor of women, young adults, and people with higher education and incomes [17]. Second, the cross-sectional design of this study limited causal inference between sources of information, worry about COVID-19, and self-confidence. Third, this study was conducted during the period of COVID-19 mitigation but not during the period when COVID-19 first emerged in Taiwan.

Conclusion

People may use a variety of sources to search for information on COVID-19, and various sources of information had various relationships with the confidence in coping with COVID-19. Medical professionals should consider subgroups of the population when establishing the various means to deliver information on COVID-19.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

C-FY and N-YK designed and conducted the study. Y-LC and D-JL analyzed the data. P-WW, C-FY, W-HL, and Y-PC drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Science and Engineering at Johns Hopkins. 2020. Modeling 2019-nCoV URL: https://systems.jhu.edu/research/public-health/ncov-model/ [accessed 2020-05-08]
2. World Health Organization. 2004. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003 URL: https://www.who.int/csr/sars/country/table2004_04_21/en/ [accessed 2020-05-08]
3. Hua J, Shaw R. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: the case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):E2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
4. National Research Council. Networking Health: Prescriptions for the Internet. Washington, DC: The National Academies Press; 2000.
5. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
6. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Apr 10:1-3 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
7. Cuan-Baltazar J, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
8. Bastani P, Bahrami M. COVID-19 related misinformation on social media: a qualitative study from Iran. J Med Internet Res 2020 Apr 05:1-5. [doi: 10.2196/18932] [Medline: 32250961]
9. Garfin D, Silver R, Holman E. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357. [doi: 10.1037/hea0000875] [Medline: 32202824]
10. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: a cross-sectional survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
11. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64. [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
12. Liao Q, Cowling B, Lam W, Ng D, Fielding R. Anxiety, worry and cognitive risk estimate in relation to protective behaviors during the 2009 influenza A/H1N1 pandemic in Hong Kong: ten cross-sectional surveys. BMC Infect Dis 2014 Mar 27;14:169 [FREE Full text] [doi: 10.1186/1471-2334-14-169] [Medline: 24674239]

13. Effective Communication in Outbreak Management for Europe. 2015. Standard questionnaire on risk perception of an infectious disease outbreak URL: http://ecomeu.info/wp-content/uploads/2015/11/Standard-questionnaire-risk-perception-ECOM-november-2015.pdf
14. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]
15. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
16. Neto MLR, Almeida HG, Esmeraldo JD, Nobre CB, Pinheiro WR, de Oliveira CRT, et al. When health professionals look death in the eye: the mental health of professionals who deal daily with the 2019 coronavirus outbreak. Psychiatry Res 2020 Jun;288:112972 [FREE Full text] [doi: 10.1016/j.psychres.2020.112972] [Medline: 32302817]
17. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]

Abbreviations

COVID-19: coronavirus disease
IRB: Institutional Review Board
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 09.05.20; peer-reviewed by MA Bahrami, K Aguirre; comments to author 02.06.20; accepted 02.06.20; published 05.06.20.

Please cite as:
Wang PW, Lu WH, Ko NY, Chen YL, Li DJ, Chang YP, Yen CF
COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan
J Med Internet Res 2020;22(6):e20021
URL: http://www.jmir.org/2020/6/e20021/
doi:10.2196/20021
PMID:32490839

Original Paper

Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States

Malcolm Fisk[1], BSc, MA, PhD; Anne Livingstone[2], MA; Sabrina Winona Pit[3,4], MSc, PhD, Dip (OHS)

[1]De Montfort University, Leicester, United Kingdom

[2]Global Community Resourcing, Brisbane, Australia

[3]University Centre for Rural Health, School of Medicine, University of Western Sydney, Lismore, Australia

[4]University Centre for Rural Health, School of Medicine and Health Sciences, University of Sydney, Lismore, Australia

Corresponding Author:
Malcolm Fisk, BSc, MA, PhD
De Montfort University
The Gateway
Leicester, LE1 9BH
United Kingdom
Phone: 44 116 257 7723
Email: malcolm.fisk@dmu.ac.uk

Abstract

Background: On March 12, 2020, the World Health Organization declared the coronavirus disease (COVID-19) outbreak a pandemic. On that date, there were 134,576 reported cases and 4981 deaths worldwide. By March 26, 2020, just 2 weeks later, reported cases had increased four-fold to 531,865, and deaths increased five-fold to 24,073. Older people are both major users of telehealth services and are more likely to die as a result of COVID-19.

Objective: This paper examines the extent that Australia, the United Kingdom, and the United States, during the 2 weeks following the pandemic announcement, sought to promote telehealth as a tool that could help identify COVID-19 among older people who may live alone, be frail, or be self-isolating, and give support to or facilitate the treatment of people who are or may be infected.

Methods: This paper reports, for the 2-week period previously mentioned and immediately prior, on activities and initiatives in the three countries taken by governments or their agencies (at national or state levels) together with publications or guidance issued by professional, trade, and charitable bodies. Different sources of information are drawn upon that point to the perceived likely benefits of telehealth in fighting the pandemic. It is not the purpose of this paper to draw together or analyze information that reflects growing knowledge about COVID-19, except where telehealth is seen as a component.

Results: The picture that emerges for the three countries, based on the sources identified, shows a number of differences. These differences center on the nature of their health services, the extent of attention given to older people (and the circumstances that can relate to them), the different geographies (notably concerned with rurality), and the changes to funding frameworks that could impact these. Common to all three countries is the value attributed to maintaining quality safeguards in the wider context of their health services but where such services are noted as sometimes having precluded significant telehealth use.

Conclusions: The COVID-19 pandemic is forcing changes and may help to establish telehealth more firmly in its aftermath. Some of the changes may not be long-lasting. However, the momentum is such that telehealth will almost certainly find a stronger place within health service frameworks for each of the three countries and is likely to have increased acceptance among both patients and health care providers.

(J Med Internet Res 2020;22(6):e19264) doi:10.2196/19264

KEYWORDS

telehealth; COVID-19; SARS-CoV-2; public health; older people; resource allocation; aged care; innovation; pandemic; telemedicine

Introduction

Context

In March 2020, the nature and virulence of the coronavirus disease (COVID-19) became a matter of urgent debate. This brought telehealth into focus as a potential tool to help provide services without the need for direct face-to-face contact. Older people, as major users of telehealth services and the age group most likely to die as a result of a COVID-19 infection, were positioned to become beneficiaries of any expanded range of these services. This may especially be the case since older people in isolation may be at risk of depression and anxiety [1,2].

The focus of this paper is on the 2-week period beginning March 12, 2020, when the World Health Organization announced that the COVID-19 outbreak was a pandemic. Notably, on this date, there was high or accelerating numbers of cases and related deaths in China (where the disease was first reported), South Korea, Iran, Italy, and Spain. Worldwide there were 134,576 reported cases and 4981 deaths. By March 26, just 2 weeks later, there were 397,289 reported cases and 24,073 deaths, both reflecting increases of over 300% [3]. These global increases were reflected in increases for each of the three countries (see Table 1).

Older people have the highest risk of mortality from COVID-19. The risk may vary per country or per region depending on various factors such as screening strategies and population distributions. The Chinese Center for Disease Control and Prevention reported that, from 72,314 cases, there was a case-fatality rate (CFR) of 8.0% for those aged 70-79 years and a 14.8% CFR for those 80 years or older [4]. The highest CFR was for people with cardiovascular disease, diabetes, chronic respiratory disease, hypertension, and cancer. Men had a higher risk of death than women (2.8% vs 1.7%). Italy has shown higher CFRs for people aged 70-79 years (19.1%) when compared to China (8.0%) [5]. Gender differences for cases and deaths were indicated for Italy and Germany, where men were, according to initial statistics, over 30% more likely to have the disease or die from it [6,7]. Such differences were noted as coming with provisos that relate to, for example, lifestyles and the extent of people's employment in relevant caring tasks [8,9]; overall (ie, for all projections at this early stage of the disease), there is further statistical uncertainty in view of data relating solely to "confirmed" cases and the omission of an "unknown" number of people who are asymptomatic [10].

Table 1. Coronavirus disease cases and deaths worldwide in Australia, the United Kingdom, and the United States from March 12 to 16, 2020.

Country	March 12, 2020		March 26, 2020		Increase	
	Cases, n	Deaths, n	Cases, n	Deaths, n	Cases, n (%)	Deaths, n (%)
Australia	156	3	3050	13	2894 (1955)	10 (433)
United Kingdom	590	10	11,658	578	11,068 (1976)	568 (5780)
United States	1697	41	85,435	1295	83,737 (5034)	1254 (3159)
World	134,576	4981	531,865	24,073	397,289 (395)	19,092 (483)

Definition of Telehealth

A note on the definition of telehealth is necessary in view of its uneasy position in relation to telemedicine, telecare, technology-enabled health, and digital health—terms that may overlap and are sometimes used interchangeably. Mobile health (mHealth; and the use of apps) are also included. Differences in understanding are indicated in the country profiles below.

The definition of telehealth offered here is encompassed by the term digital health. Its origin lies in the European Code of Practice for Telehealth Services [11]. Telehealth, it affirms, is "the means by which technologies and related services concerned with health and well-being are accessed by people or provided for them irrespective of location." This definition fits with Wootton's [12] description of telemedicine as "health care carried out at a distance," with both reflecting a person- (or patient-) centered, as opposed to a technology-driven, approach. Either can be viewed as suitable to underpin potentially new norms for health service provision, in part forced by the COVID-19 pandemic.

Benefits and Barriers of Telehealth

The reported benefits of telehealth have focused on cost, choice, and convenience. In respect to cost, much debate has taken place on the extent of financial savings that might be realized. The most substantial study of telehealth interventions, that of the "Whole System Demonstrators" (WSDs), started in 2010 and involved 3000 patients in England. It was reported that, although there was a reduction in hospital admissions, telehealth did not "seem to be a cost-effective addition to standard support and treatment" [11,13]. An ensuing study in Northern Ireland, with just under 4000 patients, noted "no evidence within the dataset of any marked impact of telehealth services on hospitalisations and hospital-based service usage" [14]. Both studies related to interventions involving the use of devices such as vital sign monitors linked to home hubs characteristic of "telecare" services that operate throughout the United Kingdom.

In contrast, an Australian study in 2013-2014 that involved nearly 300 telehealth patients explored the benefits of both vital signs monitors and technologies that included participant videoconferencing capabilities and messaging features. It, like the WSDs in the United Kingdom, found a reduction in hospital admissions but, importantly, also found a significant improvement in participants' health literacy and health behaviors, together with reported improvements in anxiety, depression, and quality of life [15,16]. The Australian study, by taking a more person-focused approach, at least *touched* on some of the potential benefits of telehealth that are not concerned with cost-effectiveness (as seen by provider organizations and funding bodies).

Other small-scale studies have taken this further, emphasizing the convenience of such services, especially when accessed by people in rural areas and, therefore, obviating the need to travel [17]. Although personal contact in the form of *presence* and *touch* is beyond telehealth, at least there is some help toward addressing the twin scourges of loneliness and social isolation that can afflict some older people. Linked with this is a real ability to make connections that give comfort and companionship to (older) people when they are dying, access to interpreters and signers where necessary, and even provide cognitive behavioral therapy.

Prominent among telehealth technologies are health apps, with their ubiquity increasing in the context of smartphone use, including messaging for mindfulness-based therapy for patients with cancer [18] or providing "digital therapy" for people with mental health needs [2]. There is an immense range of health-related apps, totaling over 300,000 in iTunes and Google Play in 2017, though the quality of most apps was considered as leaving much to be desired [19].

Known barriers to uptake of telehealth, aside from cost-effectiveness, center around the lack of (suitable) information technology and the security of communication links regarding personal (including health) data [2,20]. In addition, there are concerns regarding the impact on patient rapport, workforce, liability and legal issues, and time constraints [21]. Other barriers include the need to rethink business models and to overcome financial barriers including incentives, billing, and both initial and longer-term funding. Finally, some are concerned that telehealth is of lower efficacy; this clearly being the case where *face-to-face* contact with patients is necessary, for example, to undertake clinical examinations [17]. It is also of concern if there would be any deferring (by older people in the context of COVID-19) of making contact with health services that could assist in relation to, for example, pain, weight loss, and diminution of strength, or if any circumstances were overlooked that could relate to, for example, isolation and its consequences (eg, lack of food intake, depression, and even suicidal thoughts). Additionally, the COVID-19 outbreak could lead to a "mental health spike" [2].

Finally, there are the twin technological barriers that relate to the limitations or unreliability of internet connections (a matter that relates, in part, to the geographies of each country but also to the capacity and, therefore, the efficacy of remote connections) and the interoperability of the technologies concerned.

In relation to the positive aspects of telehealth, the need for future research has been called for by Dinesen et al [22] who affirmed the need to "identify...factors that *promote* [our emphasis] telehealth acceptance, such as human-technology interaction, organization of the health care system and social factors."

COVID-19 and Telehealth

The COVID-19 pandemic casts telehealth in a new light because it is accessed by people directly from their homes. The use of telehealth "may reduce the likelihood of viral transmission by limiting person-to-person contact, while enabling people with the virus to be treated for viral symptoms and their normal medical conditions" [23]. Health professionals meanwhile (as is already the case for extant telehealth services) are able to undertake more work from home. In addition, telehealth can be seen as safeguarding both health and social care workers, and (older) people who access such services *at least for those tasks that do not require physical contact*.

Hollander and Carr [24] pointed to the merits of telehealth enabling the diversion of people, where appropriate, from centers that deal with "emergencies" to "nurse triage lines" or for scheduled video consultations. They affirmed that remote screening and diagnosis (or referrals) can potentially reduce exposure for health care workers and other patients, as it will allow patients "to bypass the ED (Emergency Department) and be placed directly in a hospital bed." Technologies (whether via tele- or video consultations) can, they argued, inform and motivate people in relation to their lifestyles or the (self-) management of different conditions.

We can envisage, furthermore, the increased use of self-test kits with these and other vital sign measurement devices (eg, for blood pressure and respiratory function), contributing to a devolution of more traditional health care to the home. At least some of the matters that tie us to "the delivery of services in bricks-and-mortar campuses and clinics," where infection transmission is too easily facilitated, could be loosened and even make such institutions "largely unusable" [25].

It is only a small additional step to recognize the potential of telehealth through the use of artificial intelligence (AI) with, for example, remote screening via video that can recognize and record voice, facial expressions, attention, skin pallor, movement, and other signs for assessment and diagnostic purposes. However, published work in English on the use or potential use of AI in the context of both telehealth and COVID-19 is currently absent except for identifying disease outbreaks and learning from patterns of spread [26].

Particular barriers in relation to telehealth development in response to COVID-19 for three countries are noted in the Results section. Some are being addressed by changes to legislation and regulation, finance, and support programs. Protocols to guide at least video consultations were either already in place or under review. Most health and social care professionals, faced with the new demands of COVID-19, are therefore on a steep learning curve relating to the virus itself and necessary operational changes.

This paper examines the extent to which Australia, the United Kingdom (consisting of England, Scotland, Wales, and Northern Ireland), and the United States have, as a consequence of the pandemic, implemented telehealth as a tool to help older people who may live alone, are frail, or are self-isolating; identify those who may have a COVID-19 infection; and give support and facilitate treatment where necessary.

Methods

Three countries were explored: Australia, the United Kingdom, and the United States. The focus is on the 2-week period starting on March 12, 2020, when the World Health Organization announced the COVID-19 pandemic. This paper reports on the activities and initiatives taken by governments or their agencies (at national or state levels) in the three countries and the publications or guidance issued by professional, trade, and charitable bodies. Different sources of information are drawn upon that point to the perceived likely benefits of telehealth in fighting the pandemic.

Accessing the relevant sources of information for each of the countries was facilitated in large part by the knowledge and networks of the coauthors. It involved internet searches, scrutiny of media reports, and the use of contacts who work in practice or are involved in consultancy (including those who are acknowledged at the end of this paper). The focus of attention was on national (or federal) policy initiatives that would impact, through changes in funding frameworks or other strategic measures, on practice within regional, state, or territory jurisdictions.

During the period of study, the coauthors explored the burgeoning range of publications online (in the form of newsletters and blogs), some of which bore testimony to the escalation in use of tele- and video consultations, and reactions (mainly of service providers) to the funding and policy changes that were being enacted or signaled. In the wake of these a smaller but important number of academic articles were accessed—these, in general, revisited earlier work and began to re-evaluate the potential of telehealth in the new context.

The different governmental structures of the three countries, and the fragmented nature of telehealth services therein, means that they cannot be taken as representative of any wider range of countries. This must be a matter for future study.

Results

Australia

On March 11, 2020, the Australian government announced an AUS $2.4 billion (US $1.6 billion) health package to combat COVID-19. AUS $100 million (US $68 million) was promised to fund a “new Medicare service,” at no cost for patients, concerned with telehealth consultations via phone or video (eg, Skype) by general practitioners (GPs), specialists, nurses, and mental health allied health workers. The service would be available for COVID-19–related consultations and, more widely, to people at greater risk of COVID-19 infection, including those older than 70 years (or older than 50 years for Aboriginal or Torres Strait Islanders), people with chronic conditions or who are immunosuppressed, women who are pregnant, and parents with new babies. A free 24/7 national triage phone line was also to benefit from additional funding of AUS $50.7 million (US $34.4 million). In addition, AUS $25 million (US $17 million) was earmarked to fund Australians in isolation and at-risk groups to file their medication prescriptions online and have medicines home delivered free-of-charge [27].

This urgent initiative followed a call from the Royal Australian College of General Practitioners (RACGP) on March 6, 2020, for the government to “relax current restrictions around telehealth services by removing geographical constraints and permitting GPs to interact with their patients irrespective of location” [23]. Other bodies such as the Australian Medical Association, Australian College of Rural and Remote Medicine (ACRRM), and the Rural Doctors Association of Australia added their voice. Snoswell et al [23] noted that previous government telehealth funding had been made in response to droughts and bushfires, and on March 23, the Australian Government allowed vulnerable health care professionals who were authorized to use telehealth “item numbers” (ie, reimbursable) for all consultations with all their patients [28].

Australia already had relatively well-established telehealth services with some 150,000 “visits” having taken place from “rural and remote communities” in 2018 [23]. This is despite that fact that medical students are not exposed much to telehealth in their training, despite a realization of its potential benefits. In a 2018 study, they expressed preference for face-to-face consultations [17].

The ACRRM had published a standards framework (in 2016) that promoted the use of telehealth services for remote communities [29]. Guidelines for operating videoconference calls were already in place, such as from the Royal Australasian College of Physicians [30]; although, these required a GP, practice nurse, or Aboriginal health worker to be present with the patient during consultations. There remain barriers, however, to service operation in view of people having to have “access to a videoconferencing platform and internet connection” despite, as noted by Snoswell et al [23], that these could be freely or cheaply done via a “tablet or PC”. The broader context for Australia had been set in the nation’s digital health strategy that called for “widening access to telehealth services” [31].

Finally, on March 30, 2020, the Australian government announced funding of AUS $669 million (US $454 million) for the rollout of a universal telehealth model for all Australians to enable health care access through tele- or video consultations from home until September 30, 2020 [32]. There were multiple benefits: the new model enabling a reduction in COVID-19 exposure for both patients and health care providers, a maintaining of the primary health care frontline, and a reduction in the demand for personal protective equipment and emergency departments. It also helped people to stay at home, therefore, supporting compliance with self-isolation and quarantine requirements. Providers were expected to adopt either a 100% remote business model or a hybrid model of service provision. Face-to-face consultations were recognized as still being needed where physical examination was required or where technology could not be used (eg, for a confused patient without support). Practices would need to create new workflows, and some local primary health networks were guided in these tasks [33].

In addition, on March 30, 2020, further funding (AUS $74 million [US $50 million]) was provided to support telehealth consultations for those with mental health needs, including the development of a digital mental health portal and a “coronavirus hotline” for well-being and online support for health workers.

Specifically for older people, AUS $10 million (US $6.8 million) was assigned to the existing community visitors scheme (and to train volunteer visitors) to combat social isolation caused by COVID-19–imposed visiting restrictions. Such volunteer visitors will connect with older people both online and by phone.

Elsewhere, the growth in the use of apps was noted by Scott et al [19], who recognized their usefulness in relation to certain health conditions. They set out a comprehensive framework by which their merits could be assessed. An "App Evaluation Model" for mental health apps is available from the American Psychiatric Association [34]. Apps were already being trialed within some local primary health networks in Australia [35] and in the United Kingdom [36]. By March 27, 2020, a European Commission funded project had pointed to 19 mHealth "solutions" to help with the COVID-19 outbreak [37].

Given the role of apps as a means that people can self-manage and at the same time share information (eg, on heart rate and respiration) with health services, their *potential* importance as a tool of telehealth in the context of the COVID-19 pandemic is clear. Research has demonstrated that, if health care providers discuss the use of health apps with their patients, they are generally willing to use them to manage their chronic conditions [38], but such use in Australia is still in its infancy. COVID-19 might accelerate their uptake and use.

Furthermore, well-developed "aged care" services are important throughout Australia. These include the usual range of services in care homes and the wider community (home care), and have often been underpinned by "social alarms" (personal response systems) or telecare. In this sector, although not clinician-led, there is a rapidly growing awareness of telehealth's potential in supporting older people at home and of the tools that are available to help with this. The Aged Care Industry IT Council (ACIITC; that draws together Aged Care Services Australia and Leading Age Services Australia) mapped and documented recent technology changes [16]. This helped multiple aged care services to re-evaluate their roles in the context of such technologies and, where they had not already done so, to look at service cultures, operational procedures, and related training.

A leading Australian example of an *aged care* service that crosses over from the social to health care sectors is provided by Feros Care. This *aged care* service provider has partnered, for its home care services, with Google to facilitate older people's use of Google Assist, thus, giving them (and caregivers) access to the organization's portal and a widening range of information and other services. Information gathered regarding service use is envisaged as a prelude to using AI to monitor well-being [16].

The range of technologies documented in the work of the ACIITC was substantial and carries importance in the COVID-19 context. It ranged from apps and voice assistants (including Google Assist, Amazon's Alexa, and Apple's Siri) to fall-detection devices and socially assistive robots—with consideration also being given to related developments around smart homes and the internet of things. Importantly the report recognized "touch points" with clinicians. This is because of the extent that such technologies are now able to provide lifestyle and physiological data that can both help people to remain in good health and safely manage any health conditions, a matter that carries greater importance when self-isolating. It is only a small step thereafter to consider (as previously noted) how AI can be used [39] to facilitate not just monitoring (with necessary safeguards around privacy) but also diagnosis and treatment.

In the meantime, levels of awareness of the role of telehealth in relation to COVID-19 are rapidly rising. Helping this was the Digital Health Cooperative Research Centre webinar on March 18, 2020, titled "COVID-19 and Digital Technology: the Roles, Relevance and Risks of Using Telehealth in a Crisis" [40].

The resources, guidelines, training, online forums, and directory of telehealth care specialist and generalists maintained by the ACRRM may, in this context, assist health care organizations in setting up, reshaping their services, and supporting their workforces through the transitions [41]. The RACGP, furthermore, was poised to release a checklist on how to set up good clinical care in the age of telemedicine. More broadly, the Australian Digital Health Agency, having consulted on the issue of interoperability in 2019, aimed to publish a "National Health Interoperability Roadmap" [42].

However, as for all three countries, the *reach* of new telehealth initiatives to older people in Australia, despite the urgency around COVID-19, is uncertain. Many people, disproportionately those with the greatest needs, may not have (or cannot afford) smartphones or computers. Some, depending on location, have poor (or no) connectivity—albeit alleviated by the fact that many can use a landline to consult with their health care provider. Others, maybe many, may find it hard to consult over the phone and could forgo their health care *visits* until the pandemic is finished, with this potentially leading to other health care complications, the implications of which have not yet been adequately considered.

In summary, there is a good range of operational telehealth services in Australia (that offer tele- and video consultations), notably in rural areas, and such services may be able to further develop their wares as a consequence of the government's investment promise. Furthermore, the existing range of aged care services and the extent of their recognition in the role of new technologies is significant. This makes Australia relatively well positioned to respond to the COVID-19 challenge and to develop telehealth services in ways that respond to both health and social care needs.

United Kingdom

On March 12, 2020, the United Kingdom was moving to a "containment" phase in its response to the COVID-19 pandemic. The Prime Minister affirmed that, in this phase, "many more families will lose loved ones before their time" [43]. Telehealth did not have a place in the UK government's plans at this point; although Scotland had announced on March 10 that they were "accelerating" an investment of £1.24 million (US $1.5 million) plus £8 million (US $10 million) "implementation" costs to support video consultations, already used in rural areas, more widely, including for GP consultations [44].

Impetus for the UK action was added to through a publication from the Imperial College London. On March 16, 2020, Ferguson et al [45] modeled the potential of nonpharmaceutical interventions for the United Kingdom and the United States, aimed at reducing contact rates and disease transmission. It pointed to the possibility that 81% of both populations would catch the disease if control measures were not put in place. The control measures included in the modeling were "case isolation in the home," "voluntary home quarantine" of *all* household members, and "social distancing." These, of course, have a severe impact on older people.

On March 17, 2020, NHS England issued a notice to health trusts, health service commissioners (procurers), and providers, including GP services. This called for the agencies in question to "support the provision of telephone-based or digital- and video-based consultations, and advice for outpatients," and for general patient consultations to be undertaken by GPs and other health care staff. For the latter, the "roll out" of such practices would, it was considered, be accompanied by increased use of email and text messaging. "Face to face appointments," the notice stated, "should only take place when absolutely necessary."

The limited promotion of telehealth on a UK-wide basis in response to COVID-19 is likely because of the general lack of developed services (the exception being Scotland). This is despite what has been recognized as a sizeable market for such services—a major part of which was relating to mHealth and the use in the United Kingdom of apps and smart phones [46]. There are, however, many established social alarm (personal response systems) and telecare services. A leading UK example (in the north of England) of a telehealth service that crosses over from the health care to the social care sector is that provided by the Airedale NHS Foundation Trust. The "hub" provides varied services including telemonitoring, tele-coaching, the provision of advice, home visit scheduling, and (where appropriate) clinical assessments enabled through video consultations. Furthermore, is the "Gold Line" service that provides video contact for people "approaching or in the last year of life" [47,48].

The Digital Health and Care Institute, based in Scotland and financially supported by the Scottish Government, reported that some 1.8 million (mainly older) people in the United Kingdom use such services, with some also benefiting from vital sign monitoring. Most of these services, it can be noted, have their origins in housing and social care services. The TSA (formerly the Telecare Services Association) meanwhile called for their service provider members to "engage with health and social care partners" to plan for the COVID-19 response and be ready "for increased demand from vulnerable service users" [49]. It follows that many telecare services in the United Kingdom have quickly reshaped their offerings to enable staff, normally located at their monitoring and control centers, to work from home; implement new practices for visiting staff (eg, to undertake assessments or respond to urgent circumstances); and adopt, where possible, self-installation procedures (where home "hubs" are delivered for simple connection to a telephone line or internet connection).

Telehealth service development in relation to COVID-19 in the United Kingdom appears, therefore, to be initially slow. The Scottish exception, similar to Australia, will build on experience that was driven by the needs of rural, remote, and island communities. Its program includes several recognized elements of telehealth such as "home and mobile health monitoring," videoconferencing, and telecare. It also includes the use of a "bespoke" videoconferencing system for people at home (with internet-linked computers) or who are "on the move" (via smartphones). These services are not just concerned with health-related consultations but also links, at least for Scotland's most rural health service, with dispensing practices to facilitate "better pharmaceutical management" [50,51].

For the United Kingdom overall, therefore, the COVID-19 outbreak was a major "jolt" to the National Health Service (NHS) that had been and remains, in part, reluctant to embrace telehealth. A recent harbinger of necessary change was offered, however, in the Topol Review [39] that called, in the context of technological changes, for dramatic improvements in the England's health and social care infrastructure (including the workforce) and associated changes in culture. In the review, telemedicine (a subset of telehealth) topped the list of the most relevant and necessary technological advances, followed by smartphone apps and remote monitoring facilitated (in part) through sensors, including those embedded within wearable devices. Relevant also is NHSX, established in 2019 [52] as an NHS "spin-off," that is intended to lead, for England, the "largest digital health and social care transformation programme in the world," with foci that include the interoperability of systems and an intent to guide ways in which benefits for the NHS can be harnessed from "big data" analytics. Their "tech plan," including attention to apps, was under development in 2020. NHSX is, according to the NHS England notice of March 17, 2020, leading work relating to telehealth developments in the primary care sector.

In response to the question as to why telehealth had not (to date) been developed further in the United Kingdom, Professor Trish Greenhalgh put it succinctly in a webinar on March 18, 2020 [40]. She affirmed, referring to clinicians, that they "*didn't have any particular reason to use [telehealth]. They didn't see a clinical need,*" adding that "*running a service with video-consultations as a main component involves major changes in workflows and also changes in professional interactions. And it feels a bit weird* [for them] *to be consulting either by telephone or video when you could just bring the patient in and look at them - as you were taught*." Other work by Greenhalgh is relevant in exploring video consultations [53] and remote assessments [54].

Useful in this context is the release by the (UK) Royal College of General Practitioners of "top tips" for telephone consultations in the context of COVID-19 [55] and preliminary "video-consultation information" for GPs developed by the NHS in collaboration with the University of Oxford and drawing on guidance produced by the Scottish Government [56].

In summary, the *reach* of new telehealth initiatives to older people in the United Kingdom is uncertain. Many (older) people do not have smartphones or computers. Some, depending on location, will have poor (or no) connectivity. In the United Kingdom, furthermore, although there is good NHS *intent* to reach all those in need (as part of their universal service obligation), the rollout of telehealth services, despite the urgency of COVID-19, may be slow—though lessons will be learned from both Scotland and outside of the United Kingdom.

United States of America

On March 17, 2020, the United States announced the "dramatic" expansion of telehealth services via tele- and videoconferencing, with people able to use these services over the ensuing 6 months through such platforms as Skype or Facetime and with the waiving of "other" normal requirements [57]. The waiving of regulations applied to the state authorized Medicare funded services that operate throughout the United States [58] and followed growing concerns about the COVID-19 outbreak. A further US $2 trillion (of which US $200 million was earmarked for telehealth) was promised through the Coronavirus Aid, Relief, and Economic Security Act that passed through the Senate on March 25, 2020, [59] and was signed into law on March 27.

Realization of the need for such urgent action was partly prompted by an article, published on March 11, 2020, by Hollander and Carr [24]. This affirmed that, in the context of COVID-19, "direct to consumer (or on-demand) telemedicine...is both patient-centered and conducive to self-quarantine, and it protects patients, clinicians and the community from exposure." In place, furthermore, was an infrastructure in the United States that would, they envisaged, facilitate telehealth's (telemedicine's) greater use, with programs already in place for 50 of the country's state health systems.

That initial week (commencing around March 11, 2020) was marked in the United States by an "explosion of demand" that "slammed into hospitals [that were] used to delivering telehealth consults for only a handful of patients a day." Cleveland Clinic and Jefferson Health were reported as having fifteen- and twenty-fold increases, respectively, in telehealth visits in a week. In addition, Penn Medicine, because of their increased demand, "increased the number of practitioners delivering remote consults from six to 60" [25]. Further increases were expected. Additional impetus was added to the moves because of the applicability of the Imperial College London report that pointed to the equal (though far bigger in population terms) threat of deaths to the United States [45].

By March 17, 2020, the American Association of Retired Persons, the largest representative body of older people worldwide, had posted an item on their website that explained, in reassuring terms, what to expect in a virtual visit [60]. The author of the item affirmed that a virtual visit was "very similar to what would happen in person," advising people to be ready to respond to questions, "to have a pen and paper handy," and to be increasingly equipped to take physiological measures at home, notably blood pressure and temperature. The Centers for Disease Control and Prevention [61] had published interim guidance for community health staff that included the need to "explore alternatives to face-to-face triage and visits" and "identify staff to conduct telephonic and telehealth interactions," with "protocols so that staff can triage and asses patients quickly." The American Hospital Association [62] had, a year prior to the COVID-19 pandemic, called for "widespread elimination of geographical and setting locations requirements" and an expansion of "types of technology," including remote monitoring, that could be used.

The new Medicare rules began retrospectively, starting on March 6, 2020. The telehealth services in question were required to consult in "real-time" (ie, asynchronous and store and forward communication was not included), and the prior restriction to beneficiaries in rural and remote areas was removed. Reimbursement rates for service providers were set at the same rate as for face-to-face visits. Importantly, all eligible Americans became able to link to telehealth services "through video chat and online patient portals" referred to as "virtual check-ins" [59].

The range of staff engaged by service providers (at "originating" sites) was widened, overall including doctors, nurse practitioners, licensed clinical social workers, and clinical psychologists, with ordinary consultations, as well as health screening and mental health counselling, being able to be undertaken. Health staff (including nurse practitioners and physician assistants, where necessary, at "distant" sites) became eligible for payment [63]. Although the focus was on Medicare, it can be noted that several of the private insurance service providers were starting to waive costs for remote assessments and consultations, reckoning on the health benefits that would ensue [64].

What superficially may look like a consistent approach in the United States (focused mainly on videoconferencing) hides variation between the funding and administrative frameworks that operate in different states. Dinesen et al [22] reported on a "fragmented" system in the United States where the "use of technologies can create jurisdictional conflicts, policy conflicts and remain tangential to care practices rather than integrated in [the] health care infrastructure." In 2019, the Center for Connected Health Policy [65] affirmed that "no two states are alike in how telehealth is treated, despite some similarities in the language used."

A useful fact sheet for telehealth was, however, provided by the American Hospital Association [62]. This noted that, in 2017, three-quarters of US hospitals connected "patients and consulting practitioners at a distance," albeit that there were barriers to wider adoption. All states provided reimbursement under Medicaid "for some form of live video," but less than half were reimbursed for "store and forward" [62]. There are many people, furthermore, who were recognized as "low-income or uninsured" and who "may have no choice but to pay out-of-pocket for these services" [66]. Only a minority of Americans, in fact, are on Medicare, with Cahan [67] arguing that "telehealth must also reach these 281 million individuals in the under-resourced nooks and crannies of the US."

A US example of a telehealth service is provided by St Luke's Health, which operates a "virtual care center" in Boise, Idaho. The center and its staff provide video consultations and remote patient monitoring for people at other linked health care facilities, in their homes, or (via mHealth) on the move. Importantly, the service approach is seen as evolving from one that is more reactive and responsive to health needs as they arise to one that is proactive in supporting people's day to day health [68].

Finally, mention must be made of the fact that the United States has well developed personal response services (social alarms) that are frequently in place to support patients after hospital discharge or are otherwise used to enable people to contact responders after a fall or in other necessitous circumstances. Most of these are private sector services and some are seeking to evolve as telehealth services. They echo in their form the social alarm services in Australia and the United Kingdom.

As for all three countries, the *reach* of telehealth initiatives to older people in the United States, despite the urgency around COVID-19, is uncertain. Many people do not have smartphones or computers. Some people, depending on their location, will have poor (or no) connectivity. In the United States there are, furthermore, issues around the cost of services where (older) people are not eligible for Medicare.

Discussion

This paper has exposed differences in the way that Australia, the United Kingdom, and the United States responded, in the 2-week period beginning March 12, 2020, to the promotion and use of telehealth to combat the COVID-19 pandemic. These differences relate to their prior experiences with telehealth, the different health, and to some extent, social care, contexts, and cultures; the extent that specific attention was given to older people; the respective geographies (most notably in relation to rural areas); and the linked funding frameworks. Differences in understandings of definitions or the breadth of telehealth were also indicated, reflecting the nature of extant services within the countries in question.

Regarding the health context, a strong link to long-standing universal welfare models is signaled for both Australia and the United Kingdom. In the United Kingdom, a major part of telehealth and related service provision is via public and third sector (charitable or nonprofit) organizations. In Australia, provision is both from these and the private sector. The welfare element is, however, also there for the United States where Medicare and Medicare programs seek to ensure that the needs of at least some of those who are most vulnerable are met, albeit with most people normally accessing services through private providers.

All three countries give attention to older people. An interesting aspect of this, especially in Australia and the United Kingdom, are the moves toward cultural change within services, being reflected in an increased understanding of the potential of technologies to empower their users. The use of apps (notable in the United Kingdom) and voice assistant devices in the home are pointers to this, with indications in all three countries of some moves, within the context of telehealth, toward encouraging greater self-management.

The different geographies of the three countries have been observed as influential. This, in part, reflects the fact that some (often pioneer) telehealth services, were born of necessity—arising from rurality, remoteness, and island locations, and workforce shortages [69]. This is particularly the case for Australia. It is interesting to note, therefore, the actions by governments in both Australia and the United States that waived restrictions on financial support for service provision in nonrural areas. In addition, regardless of the future of telehealth, there is no doubt as to its real achieved benefits as perceived by (older) people living in rural and remote regions.

Understandings of the meaning of telehealth in the three countries were clearly influenced by the nature of services that had been established, this then being consolidated by what was determined as eligible for funding. Hence the US "model" based around video consultations can be contrasted with the less rigidly framed, but arguably more inclusive, services in Australia and the United Kingdom. The US model may, nevertheless, prove to be a useful foundation for extended services that are able to respond to the COVID-19 pandemic. By contrast, the rapid building or strengthening of telehealth services in Australia and the United Kingdom as a response to COVID-19 could prove more problematic, in the sense that greater attention will be necessary to establishing or putting in place adjustments to staffing and operational procedures able to both ensure reasonable quality standards (for technologies and services) and to respond to the certain increase in demand.

As noted in STAT [25], for the United States, "the administrative challenges are numerous. They include training doctors to deliver virtual care...there are also technology set up challenges for new users as well as a shortage of bandwidth as the internet groans under the strain of increasing use." All three countries, meanwhile, could consider the extent to which "nurse-practitioners" and "physician assistants" could play a greater part together with social care staff (and assisted by the technologies) in the operation of telehealth services. There are, in addition, all kinds of related challenges for information governance and the inevitably time-consuming tasks associated with quality testing of both the technologies and the related services. In addition, even though funding may be provided in this extraordinary time, the questions remain as to whether this will prove adequate to support the running of viable telehealth businesses in the COVID-19 context.

Common to all three countries is their commitment to at least basic service quality safeguards. These are essential and will need to be built upon if telehealth is to take its place within established health services. The COVID-19 pandemic is forcing the changes, and the question of service sustainability will, therefore, remain. We are some way from what Cahan [67] referred to as telehealth becoming "finally mainstream – overnight(ish)," and we are some way away from loosening the ties that bind us to "the delivery of services in bricks-and-mortar campuses and clinics" and face-to-face consultations and care, "where infection transmission is too easily facilitated" [25].

What begins to come through, however, is a sense of an increasing and shared recognition of how the technologies that we associate with telehealth are not only vehicles for the provision of services but also offer a means for people's empowerment. This empowerment applies to all, including older people for whom there is an imperative for equal effort to be expended in ensuring both the form of the technologies and services is appropriate to facilitate their use. Although with telehealth, the extent of personal contact in the form of *presence* and *touch* has been noted as currently out of reach; tele- or video consultations may offer a route toward helping address the twin scourges of loneliness and social isolation that can afflict some older people. Linked with this is the ability, in the context of COVID-19, of telehealth to make the links that give comfort to (older) people when they are ill, self-isolating, or being confined. In addition, for (older) people who are more mobile, there is the often-cited telehealth benefits of not needing to travel to health facilities, with the concomitant inconvenience and cost to themselves and, frequently, accompanying persons.

Further than this, the empowerment of older people living with mild dementia and their caregivers during COVID-19 has been demonstrated in Spain through television-based health and social support interventions, and the provision of telephone-based support [70]. Telehealth in the form of television-based support went, therefore, beyond an initial objective of home support and is concerned with promoting "active aging" at home. The program has enabled caregivers to provide remote support and improved cognitive function with online memory exercises.

In summary, the picture that emerges is one of uncertainties and differences for the three countries but with an increasing awareness of the part that can, and probably must, be played by telehealth in the context of COVID-19. That part will potentially have a great benefit for older people who, it has been noted, are disproportionately impacted by the virus.

However telehealth services develop in this context, there is the reassurance, of a shared concern in the three countries for at least basic quality standards to be maintained. To do this, and whether or not the immediate impact on telehealth services is sustained after the pandemic, there will be a commensurate need for telehealth (or, rather, the broader realm of digital health) to become integrated within health and social care service frameworks. In other words, telehealth must not be seen as an "alternative" form of health care. It follows that telehealth, regardless of the impact of COVID-19, must also become integrated within the training curricula for both health and social care professionals and practitioners.

Acknowledgments

The authors are grateful to Dr Kevin Doughty, T-Cubed (Bangor, Wales, UK); Kevin McSorley, Radius Connect 24 (Holywood, Northern Ireland, UK); and Karen Moore, Simple Shared Healthcare Ltd (Stafford, England, UK) for providing information regarding relevant technologies and services, and to Dr Hugh McIntyre, East Sussex Healthcare NHS Trust (Hastings, England, UK); Dr Marlene Maheu, Telebehavior Institute Inc (San Diego, CA, United States); Dr Drago Rudel, MKS Electronic Systems Ltd (Ljubljana, Slovenia); and Mark VanderWerf, Telehealth Consultant (Phoenix, AZ, United States) who provided insights and commented on earlier drafts. MF is grateful for the financial support of the Aged Care Industry Information Technology Council in his study visit to Australia during March 2020.

Authors' Contributions

MF designed the research, drafted the manuscript, and analyzed the content. SP and AL contributed to manuscript drafting and provided critical analyses of the content. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health 2020 May;5(5):e256. [doi: 10.1016/s2468-2667(20)30061-x]
2. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
3. Worldometers. 2020. COVID-19 coronavirus pandemic URL: https://www.worldometers.info/coronavirus/ [accessed 2020-04-10]
4. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
5. Statista. 2020. Coronavirus (COVID-19) death rate in Italy as of April 9, 2020, by age group URL: https://www.statista.com/statistics/1106372/coronavirus-death-rate-by-age-group-italy/ [accessed 2020-04-10]
6. Statista. Distribution of coronavirus cases in Italy as of April 9, 2020, by gender URL: https://www.statista.com/statistics/1103031/coronavirus-cases-distribution-by-gender-italy/ [accessed 2020-04-10]
7. Statista. 2020. Number of coronavirus (COVID-19) deaths in Germany in 2020, by gender and age URL: https://www.statista.com/statistics/1105512/coronavirus-covid-19-deaths-by-gender-germany/ [accessed 2020-04-10]

8. Taylor C. CNBC. 2020 Feb 18. Coronavirus is more fatal in men rather than women, major study suggests URL: https://www.cnbc.com/2020/02/18/coronavirus-is-more-fatal-in-men-than-women-major-study-suggests.html [accessed 2020-04-10]
9. Wenham C, Smith J, Morgan R, Gender and COVID-19 Working Group. COVID-19: the gendered impacts of the outbreak. Lancet 2020 Mar 14;395(10227):846-848 [FREE Full text] [doi: 10.1016/S0140-6736(20)30526-2] [Medline: 32151325]
10. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. Lancet Infect Dis 2020 Mar 12:e [FREE Full text] [doi: 10.1016/S1473-3099(20)30195-X] [Medline: 32171390]
11. Fisk M. Telehealth. In: Gu D, Dupre ME, editors. Encyclopedia of Gerontology and Population Health. Cham: Springer; 2019.
12. Wootton R. Telemedicine: fad or future? Lancet 1995 Jan 14;345(8942):73-74. [Medline: 7815881]
13. Steventon A, Bardsley M. The Nuffield Trust. 2012 Jun. The impact of telehealth on use of hospital care and mortality URL: https://www.nuffieldtrust.org.uk/files/2017-01/impact-telehealth-on-hospital-care-mortality-web-final.pdf [accessed 2020-04-10]
14. McElnay J, Clarke M, Hughes C, Agus A, Gardner E, O?Reilly D, et al. Evaluation of Past and Present Implementation of Telemonitoring NI. Belfast: Queen's University of Belfast; 2016.
15. Celler B, Varnfield M, Sparks R, Li J, Nepal S, Jang-Jaccard J, et al. Australian e-Health Research Centre (AEHRC) CSIRO. 2016. Home monitoring of chronic disease for aged care URL: https://www.csiro.au/~/media/BF/Files/Telehealth-Trial-Final-Report-May-2016_3-Final.pdf [accessed 2020-04-10]
16. Barnett K, Livingstone A, Margelis G, Tomlins G, Young R. Aged and community sector technology and innovative practice: discussion paper. Aged Care Industry Information Technology Council 2019.
17. Pit SW, Bailey J. Medical students' exposure to Knowledge and Perceptions of Telehealth Technology: Is our Future Workforce Ready to Embrace Telehealth Service Delivery? Health Education in Practice: Journal of research for professional learning 2018 Dec 17;1(2) [FREE Full text]
18. Wells C, Malins S, Clarke S, Skorodzien I, Biswas S, Sweeney T, et al. Using smart-messaging to enhance mindfulness-based cognitive therapy for cancer patients: a mixed methods proof of concept evaluation. Psychooncology 2020 Jan;29(1):212-219 [FREE Full text] [doi: 10.1002/pon.5256] [Medline: 31654533]
19. Scott IA, Scuffham P, Gupta D, Harch TM, Borchi J, Richards B. Going digital: a narrative overview of the effects, quality and utility of mobile apps in chronic disease self-management. Aust Health Rev 2020 Feb;44(1):62-82. [doi: 10.1071/AH18064] [Medline: 30419185]
20. Lynch J, Fisk M. mHealth, Privacy and the Security of Data. In: Musselwhite C, Marston H, editors. Mobile eHealth. Cham: Springer; 2017:237-249.
21. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
22. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized telehealth in the future: a global research agenda. J Med Internet Res 2016 Mar 01;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]
23. Snoswell C, Mehrotra A, Thomas A, Smith K, Haydon H, Caffery L, et al. Croakey. 2020 Mar 05. Making the most of telehealth in COVID-19 responses, and beyond URL: https://croakey.org/making-the-most-of-telehealth-in-covid-19-responses-and-beyond/ [accessed 2020-04-10]
24. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
25. Brodwin E, Ross C. STAT. 2020 Mar 17. Surge in patients overwhelms telehealth services amid coronavirus pandemic URL: https://www.statnews.com/2020/03/17/telehealth-services-overwhelmed-amid-coronavirus-pandemic/ [accessed 2020-04-10]
26. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. Lancet Digit Health 2020 Apr;2(4):e166-e167 [FREE Full text] [doi: 10.1016/S2589-7500(20)30054-6] [Medline: 32289116]
27. Prime Minister of Australia. $2.4 billion health plan to fight COVID-19 URL: https://www.pm.gov.au/media/24-billion-health-plan-fight-covid-19 [accessed 2020-04-10]
28. RACGP. RACGP. 2020 Mar 23. Expanded telehealth items for GPs URL: https://www1.racgp.org.au/newsgp/professional/expanded-telehealth-items-for-gps?feed=RACGPnewsGPArticles [accessed 2020-04-10]
29. Australian College of Rural & Remote Medicine. The ACRRM standards framework URL: http://www.ehealth.acrrm.org.au/node/2 [accessed 2020-04-10]
30. The Royal Australasian College of Physicians. 2020. Telehealth: guidelines and practical tips URL: https://www.racp.edu.au/docs/default-source/advocacy-library/telehealth-guidelines-and-practical-tips.pdf [accessed 2020-04-10]
31. Australian Digital Health Agency. Safe, seamless and secure: evolving health and care to meet the needs of modern Australia URL: https://conversation.digitalhealth.gov.au/sites/default/files/adha-strategy-doc-2ndaug_0_1.pdf [accessed 2020-05-25]
32. Prime Minister of Australia. 2020. $1.1 billion to support more mental health, medicare and domestic violence services URL: https://www.pm.gov.au/media/11-billion-support-more-mental-health-medicare-and-domestic-violence-services-0 [accessed 2020-04-10]
33. North Coast Primary Health Network. 2020 Mar 30. New North Coast telehealth and MBS resources URL: https://ncphn.org.au/archives/news/new-north-coast-telehealth-and-mbs-resources-xhp

34. American Psychiatric Association. App evaluation model URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2020-04-10]
35. South Eastern Melbourne Primary Health Network. Hello, my name is Nellie URL: https://www.semphn.org.au/resources/nellie.html [accessed 2020-04-10]
36. Schmid M. Apps: the future is mobile. In: Chambers R, Schmid M, Birch-Jones J, editors. Digital Healthcare: The Essential Guide. Oxford: Otmoor Publishing; 2016.
37. European mHealth Hub. mHealth solutions for managing the covid-19 outbreak URL: https://mhealth-hub.org/mhealth-solutions-against-covid-19 [accessed 2020-04-10]
38. Jeffrey B, Bagala M, Creighton A, Leavey T, Nicholls S, Wood C, et al. Mobile phone applications and their use in the self-management of type 2 diabetes mellitus: a qualitative study among app users and non-app users. Diabetol Metab Syndr 2019;11:84 [FREE Full text] [doi: 10.1186/s13098-019-0480-4] [Medline: 31636719]
39. Topol E. Health Education England. London; 2019. The Topol review: preparing the healthcare workforce to deliver the digital future: an independent review URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2020-05-25]
40. Digital Health CRC. 2020. COVID-19 and digital technology: the roles, relevance and risks of using telehealth in a crisis URL: http://www.digitalhealthcrc.com/telehealth-webinar/ [accessed 2020-04-10]
41. Australian College of Rural & Remote Medicine. eHealth URL: http://www.ehealth.acrrm.org.au/ [accessed 2020-04-10]
42. Australian Digital Health Agency. Better connections: your health, your say URL: https://conversation.digitalhealth.gov.au/have-your-say [accessed 2020-04-10]
43. Prime Minister's Office. 2020 Mar 12. PM statement on coronavirus: 12 March 2020 URL: https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-12-march-2020
44. Scottish Government. 2020 Mar 10. Coronavirus: speech by Cabinet Secretary for Health and Sport 10 March URL: https://www.gov.scot/publications/novel-coronavirus-covid-19-update-1/ [accessed 2020-04-10]
45. Ferguson N, Laydon D, Nedjati-Gilani G, Imai N, Baguelin M, Bhatia S, et al. Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London 2020 Mar 16;19:1-20 [FREE Full text] [doi: 10.25561/77482]
46. Rooney L, Rimpiläinen S, Morrison C, Nielsen SL. Review of emerging trends in digital health and care: a report by the Digital Health and Care Institute. University Strathclyde Glasgow 2018:e. [doi: 10.17868/67860]
47. Freed J, Lowe C, Flodgren G, Binks R, Doughty K, Kolsi J. Telemedicine: is it really worth it? A perspective from evidence and experience. J Innov Health Inform 2018 Mar 15;25(1):14-18. [doi: 10.14236/jhi.v25i1.957] [Medline: 29717950]
48. NHS Providers. 2015 Aug. Telemedicine at Airedale NHS Foundation Trust: better care in the community for elderly patients URL: http://www.airedaledigitalcare.nhs.uk/seecmsfile/?id=33 [accessed 2020-05-25]
49. TSA. The voice of Technology Enabled Care URL: https://www.tsa-voice.org.uk/ [accessed 2020-04-10]
50. Technology Enabled Care Programme: Data Review and Evaluation Options Study, Summary Report. Edinburgh: Scottish Government; May 2018.
51. Pike H. The Pharmaceutical Journal. 2018 Aug 09. Web therapy: how pharmacists are leading the way on telehealth URL: https://www.pharmaceutical-journal.com/news-and-analysis/features/web-therapy-how-pharmacists-are-leading-the-way-on-telehealth/20205262.article [accessed 2020-05-25]
52. NHSx. Driving forward the digital transformation of health and social care URL: https://www.nhsx.nhs.uk/ [accessed 2020-04-10]
53. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
54. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
55. Stockley S, Neighbour R. Royal College of General Practitioners. 2020 Mar 19. Top tips for COVID-19 telephone consultations URL: https://www.rcgp.org.uk/about-us/rcgp-blog/top-10-tips-for-covid-19-telephone-consultations.aspx [accessed 2020-04-10]
56. University of Oxford. 2020. Video consultations: information for GPs URL: https://bjgp.org/sites/default/files/advanced-pages/20Mar_COVID_VideoConsultations.pdf [accessed 2020-04-10]
57. Global News. 2020. Coronavirus outbreak: Trump announces expansion of medicare telehealth services amid pandemic URL: https://globalnews.ca/video/6689942/coronavirus-outbreak-trump-announces-expansion-of-medicare-telehealth-services-amid-pandemic [accessed 2020-04-10]
58. Centers for Medicare & Medicaid Services. 2020 Mar 17. President Trump expands telehealth benefits for medicare beneficiaries during COVID-19 outbreak URL: https://www.cms.gov/newsroom/press-releases/president-trump-expands-telehealth-benefits-medicare-beneficiaries-during-covid-19-outbreak [accessed 2020-04-10]
59. Sprunt B. NPR. 2020 Mar 25. Read: $2 trillion coronavirus relief bill URL: https://www.npr.org/2020/03/25/820759545/read-2-trillion-coronavirus-relief-bill?t=1585296158411 [accessed 2020-04-10]
60. Levine H. AARP. 2020 Mar 17. The benefit of telehealth during the Pandemic URL: https://www.aarp.org/health/conditions-treatments/info-2020/benefits-telehealth-medicare.html [accessed 2020-04-10]

61. Centers for Disease Control and Prevention. 2020. Healthcare facilities: preparing for community transmission URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-hcf.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fhealthcare-facilities%2Fguidance-hcf.html
62. American Hospital Association. 2019 Feb. Fact sheet: telehealth URL: https://www.aha.org/system/files/2019-02/fact-sheet-telehealth-2-4-19.pdf [accessed 2020-05-25]
63. Department of Health and Human Services. 2020 Mar 19. Notification of enforcement discretion for telehealth remote communications during the COVID-19 nationwide public health emergency URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html [accessed 2020-05-25]
64. Gorke J. Forbes. 2020 Apr 02. COVID-19 and telehealth: crisis drives flexibility and expands care URL: https://www.forbes.com/sites/jeffgorke/2020/04/02/covid-19-and-telehealth-crisis-drives-flexibility-and-expands-care/ [accessed 2020-05-25]
65. Center for Connected Health Policy. 2020. State telehealth laws and reimbursement policies report URL: https://www.cchpca.org/sites/default/files/2020-05/CCHP_%2050_STATE_REPORT_SPRING_2020_FINAL.pdf
66. Khalid A. Quartz. 2020 Mar 20. Telemedicine struggles to be an option for everyone in the wake of coronavirus URL: https://qz.com/1821549/telemedicine-faces-unprecedented-demand-in-the-wake-of-coronavirus/
67. Cahan E. TechCrunch. 2020 Apr 04. Why telehealth can't significantly flatten the coronavirus curve-yet URL: https://techcrunch.com/2020/04/04/why-telehealth-cant-significantly-flatten-the-coronavirus-curve-yet/ [accessed 2020-05-25]
68. Wicklund E. mHealth Intelligence. 2018 Aug 28. New telemedicine center to extend connected health across Idaho URL: https://mhealthintelligence.com/news/new-telemedicine-center-to-extend-connected-health-across-idaho [accessed 2020-05-25]
69. Bradford NK, Caffery LJ, Smith AC. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. Rural Remote Health 2016;16(4):3808 [FREE Full text] [Medline: 27744708]
70. Goodman-Casanova JM, Dura-Perez E, Guzman-Parra J, Cuesta-Vargas A, Mayoral-Cleries F. Telehealth home support during COVID-19 confinement for community-dwelling older adults with mild cognitive impairment or mild dementia: survey study. J Med Internet Res 2020 May 22;22(5):e19434 [FREE Full text] [doi: 10.2196/19434] [Medline: 32401215]

Abbreviations

ACRRM: Australian College of Rural and Remote Medicine
ACIITC: Aged Care Industry IT Council
AI: artificial intelligence
CFR: case-fatality rate
COVID-19: coronavirus disease
GP: general practitioner
mHealth: mobile health
NHS: National Health Service
RACGP: Royal Australian College of General Practitioners
WSD: Whole System Demonstrator

Edited by G Eysenbach; submitted 10.04.20; peer-reviewed by SY Jung, S Omboni, C Wu; comments to author 23.05.20; revised version received 24.05.20; accepted 27.05.20; published 09.06.20.

Please cite as:
Fisk M, Livingstone A, Pit SW
Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States
J Med Internet Res 2020;22(6):e19264
URL: http://www.jmir.org/2020/6/e19264/
doi:10.2196/19264
PMID:32463377

Original Paper

Development of an openEHR Template for COVID-19 Based on Clinical Guidelines

Mengyang Li[1,2]; Heather Leslie[3,4], BA; Bin Qi[5], BA; Shan Nan[1,2], PhD; Hongshuo Feng[1,2]; Hailing Cai[1,2], BA; Xudong Lu[1,2,6], PhD; Huilong Duan[1,2], PhD

[1]College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China
[2]Key Laboratory for Biomedical Engineering, Ministry of Education, Hangzhou, Zhejiang, China
[3]Atomica Informatics, Melbourne, Australia
[4]openEHR Foundation, London, United Kingdom
[5]Hangzhou Joyrun Medical Technology Cooperation, Hangzhou, Zhejiang, China
[6]School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Xudong Lu, PhD
College of Biomedical Engineering and Instrument Science
Zhejiang University
Room 512, Zhouyiqing Building
38 Zheda Road
Hangzhou, Zhejiang, 310007
China
Phone: 86 13957118891
Email: lvxd@zju.edu.cn

Abstract

Background: The coronavirus disease (COVID-19) was discovered in China in December 2019. It has developed into a threatening international public health emergency. With the exception of China, the number of cases continues to increase worldwide. A number of studies about disease diagnosis and treatment have been carried out, and many clinically proven effective results have been achieved. Although information technology can improve the transferring of such knowledge to clinical practice rapidly, data interoperability is still a challenge due to the heterogeneous nature of hospital information systems. This issue becomes even more serious if the knowledge for diagnosis and treatment is updated rapidly as is the case for COVID-19. An open, semantic-sharing, and collaborative-information modeling framework is needed to rapidly develop a shared data model for exchanging data among systems. openEHR is such a framework and is supported by many open software packages that help to promote information sharing and interoperability.

Objective: This study aims to develop a shared data model based on the openEHR modeling approach to improve the interoperability among systems for the diagnosis and treatment of COVID-19.

Methods: The latest Guideline of COVID-19 Diagnosis and Treatment in China was selected as the knowledge source for modeling. First, the guideline was analyzed and the data items used for diagnosis and treatment, and management were extracted. Second, the data items were classified and further organized into domain concepts with a mind map. Third, searching was executed in the international openEHR Clinical Knowledge Manager (CKM) to find the existing archetypes that could represent the concepts. New archetypes were developed for those concepts that could not be found. Fourth, these archetypes were further organized into a template using Ocean Template Editor. Fifth, a test case of data exchanging between the clinical data repository and clinical decision support system based on the template was conducted to verify the feasibility of the study.

Results: A total of 203 data items were extracted from the guideline in China, and 16 domain concepts (16 leaf nodes in the mind map) were organized. There were 22 archetypes used to develop the template for all data items extracted from the guideline. All of them could be found in the CKM and reused directly. The archetypes and templates were reviewed and finally released in a public project within the CKM. The test case showed that the template can facilitate the data exchange and meet the requirements of decision support.

Conclusions: This study has developed the openEHR template for COVID-19 based on the latest guideline from China using openEHR modeling methodology. It represented the capability of the methodology for rapidly modeling and sharing knowledge through reusing the existing archetypes, which is especially useful in a new and fast-changing area such as with COVID-19.

(J Med Internet Res 2020;22(6):e20239) doi:10.2196/20239

KEYWORDS

coronavirus disease; COVID-19; openEHR; archetype; template; knowledge modeling; clinical guidelines

Introduction

The coronavirus disease (COVID-19) is a severe infectious disease that has been confirmed to lead to human-to-human transmission since December 2019 [1]. Considering the sudden outbreak of the disease as a challenging threat, it was brought into the Class B of infectious diseases defined in the Law of the People's Republic of China on the Prevention and Control of Infectious Diseases, but management and policies were adopted according to the Class A of infectious diseases [2,3]. All of the challenges have brought great pressure on medical institutions and professionals, including the lack of medical equipment and the complexity of diagnosis. At the same time, the fear of the disease has a negative impact on both the psychological and physiological well-being of affected individuals.

Although, most cases of the disease occurred in mainland China in the beginning, other areas have also confirmed cases of the same disease, and the number of cases continues to increase. The World Health Organization (WHO) has declared the 2019-20 coronavirus outbreak to be a Public Health Emergency of International Concern [4]. So far, the United States is the most seriously affected country in the world. In addition to China, many cases were found in the Western Pacific Region, such as South Korea and Japan. Russia and Great Britain were relatively seriously affected areas in Europe [5] according to the latest data reported by the WHO.

Symptoms include fever, cough, or shortness of breath, and even pneumonia, multi-organ failure, and death in the most severe cases. The latent period can be between 1 and 14 days, and on average is between 3 and 7 days according to the epidemiological investigation. What is even worse is that some patients may be asymptomatic at the beginning, which results in some undetected errors [6]. Given the severity of the infectious disease and urgency of diagnosis and treatment, a large number of studies related to disease prevention and control have been carried out with the support of various countries according to the WHO [7].

Considering the rapid spread of the disease, the transferring of knowledge of diagnosis and treatment of the disease and newly updated achievements of research are important, especially from areas with improved epidemics like mainland China, the area where the epidemic had begun. Although some efforts have been made through teleconsultation and medical staff assistance, they are still limited due to a lack of experts. Using decision support tools is an efficient way to transfer the knowledge of experts to clinical practices. Epic (Epic Systems), which is a health care software company with electronic medical record software application, has sent out an update to its customers to detect potential cases of COVID-19 [8]. DIPS, which is a major openEHR vendor, released open source components to assist software developers creating apps to fight COVID-19 [9]. Although Epic provided a complete solution for medical information systems where the clinical decision support tools have already been embedded, most other systems still need to integrate with medical information systems to be used in clinical practice. The interoperability of disease-related data has become an important issue.

Many studies on terminology standardization have been conducted to improve the interoperability. Systematized Nomenclature of Human Medicine International has issued an interim release to promote the analysis with the most up-to-date terminology [10]. The Observational Health Data Sciences and Informatics has also committed vocabulary about COVID-19 into GitHub [11]. These efforts have been focused on the shared representation of the concepts related to COVID-19 but are not enough for data exchanging, with the shared data model being the key issue. It specifies not only the data structures but also the attributes of data elements. Although there exist many methods [12-15] to develop a data model for certain requirements, the rapidly updating knowledge for COVID-19 makes it a challenge to achieve a flexible model constantly with the requirement evolution.

An open, semantic-sharing, and collaborative-modeling framework is needed to meet the dynamic change of data requirements. openEHR specifications can be used to create standards and build information and interoperability solutions for health care as a multilevel modeling framework [16]. In the approach suggested by openEHR, the reference model (RM) focuses on the logic structures and attributes required to express data, so it is stable and provides basic components for building concrete medical information models. The archetype model is comprised of archetypes and templates. Based on the RM, archetypes can be developed to define all the attributes about specific clinical concepts. Different archetypes can be organized into context-specific data sets, templates that are mostly developed and used locally. Only the RM is implemented in apps, while clinical information (archetypes and templates) is independent of specific implementations. The approach allows data models to be flexible and extensible within the constraints of the RM, which can keep up with the development of the clinical knowledge and meet the requirements of the complicated clinical environment [17]. The knowledge obtained in clinical practice in mainland China can be sharable and beneficial for other countries and regions by being formalized as openEHR archetypes and templates.

As a goal, we developed an openEHR template to promote interoperability among clinical systems for the diagnosis and treatment of COVID-19. The remainder of this paper is organized as follows: the Methods section introduces the knowledge source and methodology we used to develop and review the template, the Results section illustrates our results step-by-step according to the proposed methodology, and the Discussion section discusses the contributions of this paper and limitations.

Methods

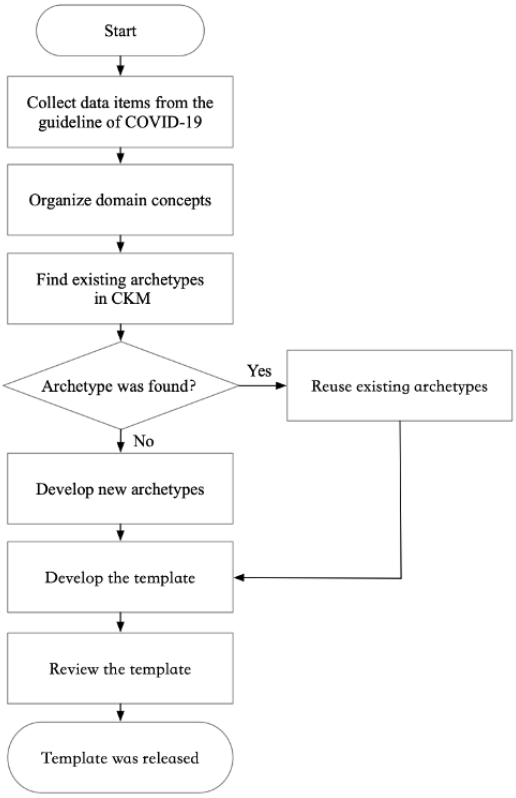
Knowledge Sources

Given that the outbreak of the disease happened within a short time frame, the involved knowledge is limited. To make them justified and believed, the Guideline for Diagnosis and Treatment of COVID-19 released by the National Health Commission of the People's Republic of China was adopted as the knowledge source. At present, the guideline has evolved to the 7th edition [18] (the English translated version is also available [19]). The guideline can be divided into 13 sections: (1) Pathogenic Characteristics; (2) Epidemiological Characteristics; (3) Pathological Changes; (4) Clinical Characteristics; (5) Diagnostic Criteria; (6) Clinical Classifications; (7) Clinical Warning Signs for Severe and Critical Cases; (8) Differential Diagnosis; (9) Discovery and Reporting of Cases; (10) Treatment; (11) Criteria for Discharge and Notes After Discharge; (12) Transportation Principles; and (13) Prevention of Infection in Medical Establishments. Since only the contents related to conditions and recommendations for diagnosis and treatment will be used for data exchanging, sections 4-9, section 11, and part of section 10 were selected as the knowledge source. There are three reasons for excluding other sections. First, sections 1-3 describe the general knowledge of COVID-19, most of which will not be directly used as the condition for the judgment of diagnosis, treatment, and management. Second, sections 12 and 13 illustrate the regulations and policies for transportation and infection prevention, and will also not be used for diagnosis and treatment. Third, in another part of section 10, traditional Chinese medicine and herbal medicine are regarded as a kind of alternative medicine, which is a supplement to evidence-based medicine and has only regional characteristics.

The Process of Development

To develop an openEHR template for COVID-19, our method consists of six steps. These steps include collecting data items, organizing domain concepts, searching corresponding archetypes, developing an openEHR template, and reviewing and releasing the template (see Figure 1).

Figure 1. The method of developing an openEHR template about COVID-19. CKM: Clinical Knowledge Manager; COVID-19: coronavirus disease.



Collecting Data Items

In this step, data items related to diagnosis and treatment were extracted from sections 4-9, section 11, and part of section 10 of the guideline, and further organized in Excel (Microsoft Corporation) with three columns. The first and the second column corresponded to the sections and subsections of the guideline, and the third column corresponded to the data items extracted from the subsection. The extracted original data items in Chinese have been translated from Chinese to English. Although extraction of data items was done manually, two principles were followed to lower the bias of extraction and reduce the errors.

First, considering that the composition of the guideline is hierarchical and segmented, which is the inherent knowledge of grouping data items, the extracted data items were organized in the same hierarchical structure as the guideline to not only lay the foundation for further organization of domain concepts but also provide a much easier correspondence for reviewers when verifying the correctness of the extraction.

Second, two members of our team extracted these data items separately. After the extraction, both of them exchanged and reviewed the opponent's results. For the results acknowledged by both of them, they were included in the Excel file directly. For the results acknowledged by only one of them, they were reviewed by another member to confirm the final results. For the results that were acknowledged by both of them but needed to be refined, they were re-extracted by both team members.

Organization of Domain Concepts

The organization of domain concepts is the basis for the development of archetypes and templates. Five steps were performed to organize domain concepts from the extracted data items.

1. If the data items from different subsections were the same semantically, they were merged into a single-data item. For example, coagulopathy and blood coagulation disorder can be merged into blood coagulation disorder.
2. If the data items from different subsections belong to the same domain concept, they were regrouped into a more suitable group other than the sections or subsections. For example, symptoms such as fever and difficulties in breathing found in different sections will be regrouped together.
3. According to practices from clinical decision support with the clinician participants, medical concepts that are encountered and used commonly were selected and organized as a supplement. For example, in diagnosis and treatment, operations such as surgery are generally mentioned, such as pneumonectomy or splenectomy, but similar medical concepts are not mentioned in the guideline, so this step is a significant supplement to the knowledge extracted from the guideline.
4. All domain concepts were then organized as a tree structure according to the inherent correlation among them and were represented into a mind map using XMind (XMind Ltd) as a tool. The extracted data items can be either the data elements themselves or one value in the value set of the data elements within domain concepts. For example, respiratory failure and blood coagulation disorder are each treated as a single value within the value set of diagnoses.
5. Finally, the domain concepts were further classified into three categories according to the different stages in the process of clinical diagnosis and treatment; they are "Instruction, Evaluation, and Observation."

Searching in Clinical Knowledge Manager and Archetype Development

To avoid developing archetypes repeatedly and to facilitate semantic interoperability, the adoption of existing archetypes is of much significance. The openEHR Foundation provides a website called Clinical Knowledge Manager (CKM) [20], which supports international domain knowledge governance and collaborative development of clinical knowledge resources beyond a library of openEHR archetypes and templates.

This step mainly focused on performing a search in the repository to find the archetypes with similar semantics. The name of domain concepts and data items were used as keywords to identify the archetypes. On account of polysemy and synonym, extra manual work was carried out to find the related archetypes. Some archetypes can be used directly, which means the data elements can be represented in these archetypes exactly, and there is no difference on a semantic level among them.

If no corresponding archetypes exist or existing archetypes cannot represent the data elements fully, developing new archetypes or extending existing archetypes is necessary according to the syntax of openEHR Archetype Definition Language [21].

Development of openEHR Template

After the required archetypes were found and developed, the template can be built based on them. This task can be performed with the support of Ocean Template Editor [22] to set constraints on these archetypes to fit the requirements for data exchanging. In this step, two issues come in to focus. First, constraints about terminology for certain data elements in the archetype should be made, such as medication. For example, the name of drugs can be generic names that can be identified around the world and aliases, which can be used in specific countries or regions. A unified value domain can be significant for data sharing and interoperability. Second, from the perspective of diagnosis and treatment, time series problems should be taken into consideration. For example, "patients whose chest imaging shows a significant progression of lesions (>50%) are managed as a severe case within 24-48 hours" was mentioned in the guideline. To support the diagnosis decision, the occurrence of the data elements in the template should be more than one so that they can contain the required two elements including the first chest imaging and the second chest imaging within 24-48 hours.

Review and Release of the Template

The review process is necessary to achieve a template with high quality. Two aspects of the template have been reviewed. First, the representation of the domain knowledge, such as correctness of the semantics, classification of data elements, and logic structure of archetypes, was reviewed. Second, the template was reviewed from informatics, such as the data types of data elements and relationships among different archetypes. The study has designed two review phases to achieve the goal.

1. Internal review phase: the internal review group includes a total of four persons, with one person that is familiar with the COVID-19 guideline, two persons who are developers of clinical data repositories (CDRs) and decision support tools, and one person that is familiar with openEHR specifications.
2. Outreach review phase: the participation of two domain experts beyond the research team was involved in this

phase. One has the expertise of openEHR modeling and the other has the expertise of medical informatics.

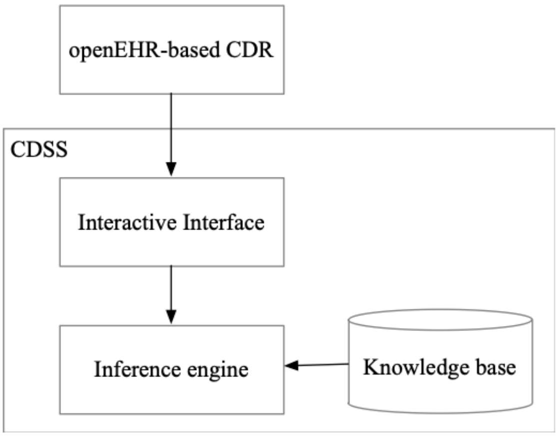
After the review, the template and the used archetypes have been uploaded and shared in the Healthcare Modeling Collaboration [23] and in the project of the CKM.

Verification of the openEHR Template COVID-19

The study has designed a test case to verify the feasibility of the template. The case was conducted in a hospital located in Wuhan, which has already implemented openEHR-based CDR and accepted a large number of patients with COVID-19. The CDR in the hospital was built based on the solution we have proposed, which can be found in [24,25]. With a developed openEHR model, the storage structure can be generated easily. The decision support tool for diagnosis and treatment of COVID-19 has been developed and is planned to be used in the practice. There is great demand for data sharing and interoperability between CDR and the decision support tool.

The test case was designed to include two steps: (1) the template was applied to the openEHR-based CDR, which can provide template-specific data query and storage service through application programming interfaces (APIs); and (2) the decision support tool used these APIs to query the data useful for the judgment of the diagnosis and treatment. The interaction diagram is shown in Figure 2.

Figure 2. The interaction diagram between CDR and decision support tool. CDR: clinical data repository; CDSS: clinical decision support system.



Results

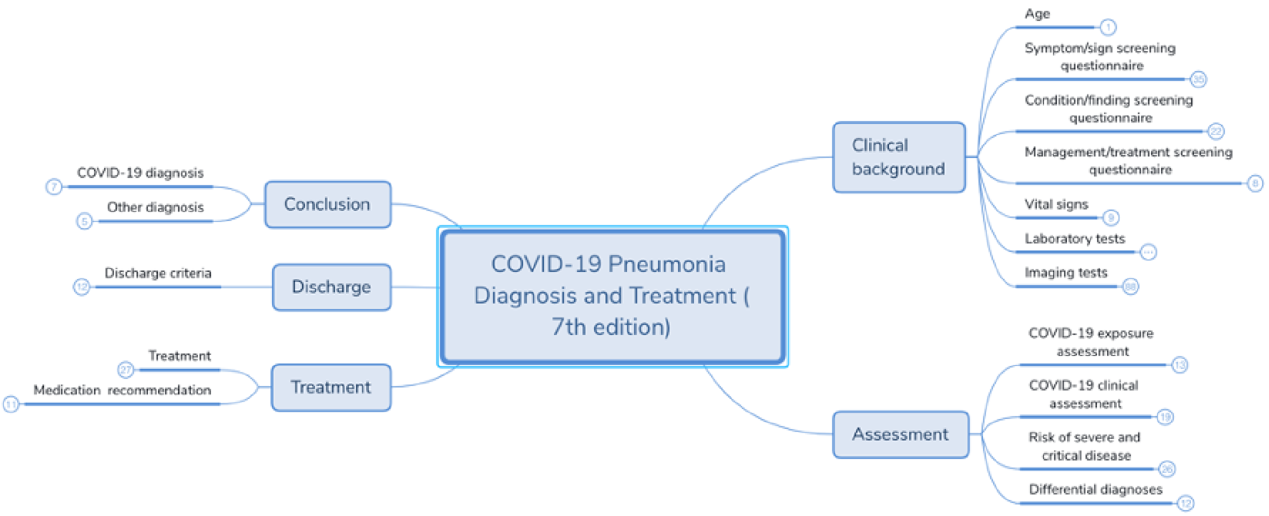
Based on the methodology previously described, 203 data items were extracted from the guideline in China, including 8 sections and 15 subsections (see Multimedia Appendix 1). After the classification and merge of these data elements, 16 domain concepts (16 leaf nodes in the mind map) were organized for diagnosis and treatment of COVID-19. The results in this step are illustrated in the mind map in Figure 3 (full results can be found in Multimedia Appendix 2).

Among these domain concepts, only 2 archetypes were classified into Instruction, and 3 of them were classified into Evaluation. The archetypes of Observation include 11 items. A total 22 archetypes have been developed to represent all data elements about COVID-19, and all of them can be referred to from the CKM directly. These archetypes found in the CKM, which are adapted to our requirements, are shown in Textbox 1. Finally, a template was developed with the constraint of these archetypes as shown in Figure 4.

In addition, it has been deployed in a hospital, which has accepted many cases of COVID-19, to support data sharing between CDRs and clinical decision support systems (CDSS). Because the CDR is developed based on openEHR, the storage structure is consistent with the template. Although there exist many storage implementations of openEHR [25,26], they are transparent for the invocation of data services as a result of openEHR two-level modeling. Data elements in the template can be uniquely identified by paths and attributes. In this way, data for each data element can be transferred in representational state transfer (REST)ful API. The rules in CDSS were also built based on the template. The template was used in both CDR and CDSS. The data from RESTful API was parsed, extracted, and used in the inference engine to produce the decision for the case. The user interface for the data view whose data was from openEHR-based CDRs is shown in Figure 5. Textbox 2 shows the part of the content of the data that is in the RESTful API from CDR.

In the end, the template (COVID-19 Pneumonia Diagnosis and Treatment [7th edition]) has been uploaded into the CKM [27].

Figure 3. Domain concepts about COVID-19. COVID-19: coronavirus disease.



Textbox 1. Domain concepts and their archetypes found in the Clinical Knowledge Manager.

Diagnosis and treatment of the coronavirus disease (COVID-19)

- openEHR-EHR-COMPOSITION.encounter.v1

—

- openEHR-EHR-SECTION.adhoc.v1

Age

- openEHR-EHR-OBSERVATION.age.v0

Symptom and sign screening questionnaire

- openEHR-EHR-OBSERVATION.symptom_sign_screening.v0

Condition and findings screening questionnaire

- openEHR-EHR-OBSERVATION.condition_screening.v0

Management and treatment screening questionnaire

- openEHR-EHR-OBSERVATION.management_screening.v0

Vital signs

- openEHR-EHR-OBSERVATION.body_temperature.v2
- openEHR-EHR-OBSERVATION.respiration.v2
- openEHR-EHR-OBSERVATION.pulse_oximetry.v1

Laboratory tests

- openEHR-EHR-OBSERVATION.laboratory_test_result.v1
- openEHR-EHR-CLUSTER.specimen.v1
- openEHR-EHR-CLUSTER.laboratory_test_analyte.v1
- openEHR-EHR-CLUSTER.inspired_oxygen.v1
- openEHR-EHR-OBSERVATION.pf_ratio.v0

Imaging tests

- openEHR-EHR-OBSERVATION.imaging_exam_result.v0
- openEHR-EHR-CLUSTER.imaging_finding.v0

COVID-19 exposure assessment

- openEHR-EHR-OBSERVATION.exposure_assessment.v0

COVID-19 clinical assessment

- openEHR-EHR-EVALUATION.health_risk.v1

Risk of severe and critical disease

- openEHR-EHR-EVALUATION.health_risk.v1

Differential diagnoses

- openEHR-EHR-EVALUATION.differential_diagnoses.v0

Treatment

- openEHR-EHR-INSTRUCTION.therapeutic_activity_order.v0

Medication recommendation

- openEHR-EHR-INSTRUCTION.medication_order.v2

Discharge criteria

- openEHR-EHR-EVALUATION.health_risk.v1

COVID-19 diagnosis

- openEHR-EHR-EVALUATION.problem_diagnosis.v1

Other diagnosis

- openEHR-EHR-EVALUATION.problem_diagnosis.v1

Figure 4. The developed template in Ocean Template Editor. COVID-19: coronavirus disease; CT: computed tomography; ESR: erythrocyte sedimentation rate; SARS: severe acute respiratory syndrome.

```
Template Properties
COVID-19 Pneumonia Diagnosis and Treatment (7th edition)   Δ NAME (from 'Encounter')
  context
  content
    Clinical background   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
      items
        Age
        Symptom/sign screening questionnaire
        Condition/finding screening questionnaire   Δ [0..*] to [0..1], NAME (from 'Condition screening questionnaire')
        Management/treatment screening questionnaire
        Vital signs   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
          items
            Body temperature
            Respiration
            Pulse oximetry
        Laboratory tests   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
          items
            SARS coronavirus 2 RNA   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Serology   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Complete blood count & differential   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Arterial blood gases   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            C reactive protein   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Interleukin 6   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Liver function panel   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Creatine kinase   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Troponins   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Myoglobin   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            ESR   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Fibrin D-dimer   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Lactate   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Lactate dehydrogenase   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            Procalcitonin   Δ [0..*] to [0..1], NAME (from '*Laboratory test result(en)')
            PaO2/FiO2 ratio
        Imaging tests   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
          items
            Chest x-ray   Δ [0..*] to [0..1], NAME (from 'Imaging examination result')
            CT Chest   Δ [0..*] to [0..1], NAME (from 'Imaging examination result')
    Assessment   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
      items
        COVID-19 exposure assessment   Δ [0..*] to [0..1], NAME (from 'Exposure assessment')
        COVID-19 clinical assessment   Δ [0..*] to [0..1], NAME (from 'Health risk assessment')
        Risk of severe and critical disease   Δ [0..*] to [0..1], NAME (from 'Health risk assessment')
        Differential diagnoses
    Treatment   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
      items
        Treatment   Δ [0..*] to [0..1], NAME (from 'Therapeutic activity')
        Medication recommendation   Δ [0..*] to [0..1], NAME (from 'Medication order')
    Discharge   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
      items
        Discharge criteria   Δ [0..*] to [0..1], NAME (from 'Health risk assessment')
    Conclusion   Δ [0..*] to [0..1], NAME (from 'Ad hoc heading')
      items
        COVID-19 diagnosis   Δ [0..*] to [0..1], NAME (from 'Problem/Diagnosis')
        Other diagnosis   Δ [0..*] to [0..1], NAME (from 'Problem/Diagnosis')
```

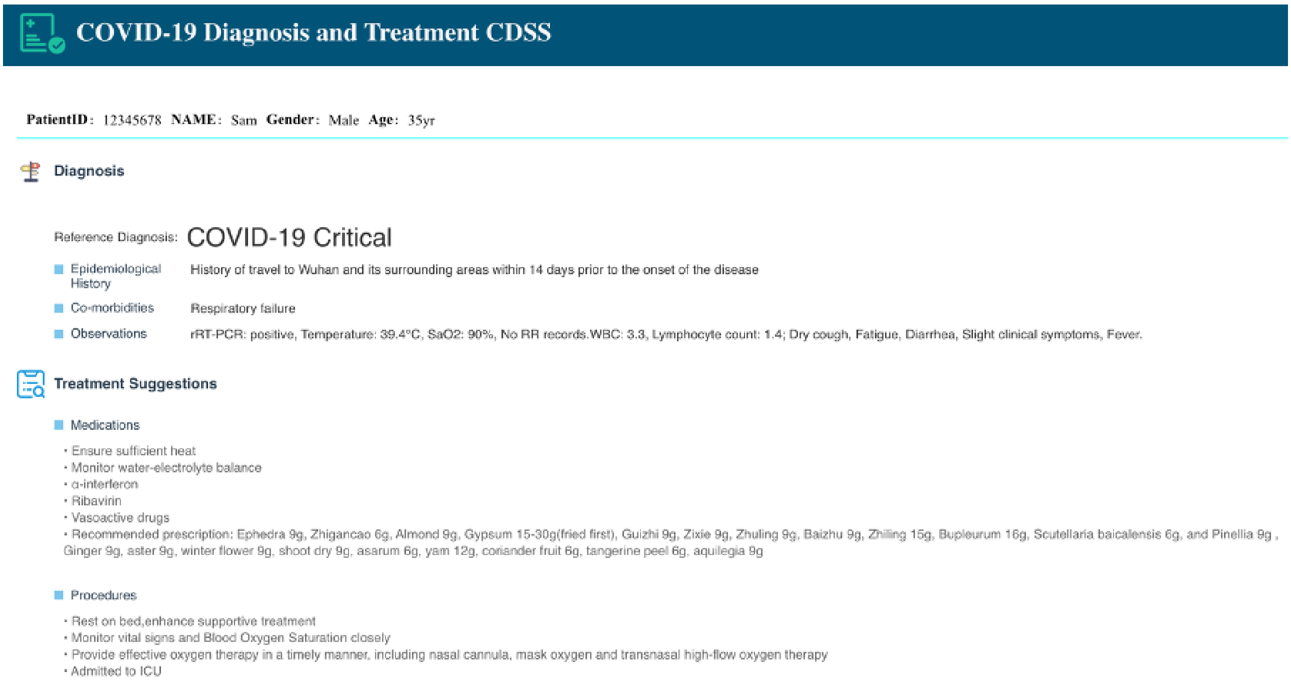


Figure 5. The data view of COVID-19 Diagnosis and Treatment CDSS. CDSS: clinical decision support system; COVID-19: coronavirus disease; ICU: intensive care unit; rRT-PCR: real time reverse transcription-polymerase chain reaction; WBC: white blood cell.

COVID-19 Diagnosis and Treatment CDSS

PatientID: 12345678 **NAME:** Sam **Gender:** Male **Age:** 35yr

Diagnosis

Reference Diagnosis: COVID-19 Critical

Epidemiological History	History of travel to Wuhan and its surrounding areas within 14 days prior to the onset of the disease
Co-morbidities	Respiratory failure
Observations	rRT-PCR: positive, Temperature: 39.4°C, SaO2: 90%, No RR records.WBC: 3.3, Lymphocyte count: 1.4; Dry cough, Fatigue, Diarrhea, Slight clinical symptoms, Fever.

Treatment Suggestions

Medications

- Ensure sufficient heat
- Monitor water-electrolyte balance
- α-interferon
- Ribavirin
- Vasoactive drugs
- Recommended prescription: Ephedra 9g, Zhigancao 6g, Almond 9g, Gypsum 15-30g(fried first), Guizhi 9g, Zixie 9g, Zhuling 9g, Baizhu 9g, Zhiling 15g, Bupleurum 16g, Scutellaria baicalensis 6g, and Pinellia 9g , Ginger 9g, aster 9g, winter flower 9g, shoot dry 9g, asarum 6g, yam 12g, coriander fruit 6g, tangerine peel 6g, aquilegia 9g

Procedures

- Rest on bed,enhance supportive treatment
- Monitor vital signs and Blood Oxygen Saturation closely
- Provide effective oxygen therapy in a timely manner, including nasal cannula, mask oxygen and transnasal high-flow oxygen therapy
- Admitted to ICU

Textbox 2. The part of the content of data in representational state transfer application programming interface from clinical data repositories.

```
{
   “labTestResultList”: [
        {
              “itemName”: “Lymphocyte count”,
              “result”: 1.4,
              “status”: “”,
              “unit”: “”
        },
        {
              “itemName”: “WBC”,
              “result”: 3.3,
              “status”: “”,
              “unit”: “”
        },
        {
              “itemName”: “rRT-PCR”,
              “result”: “positive”,
              “status”: “”,
              “unit”: “”
        }
    ],
    “medicalRecordList”: [
        {
              “dateTime”: “”,
              “text”: “History of travel to Wuhan and its surrounding areas within 14 days prior to the onset of the disease”,
              “type”: “Epidemic History”
        }
    ],
    “patientInfo”: {
        “bloodType”: [],
        “dateOfBirth”: “1985-01-01”,
        “name”: “Sam”,
        “patientId”: 12345678,
        “sex”: “Male”
    },
    “physicalSignList”: [
        {
              “itemCode”: “”,
              “itemName”: “SpO2”,
              “measureDateTime”: “”,
              “unit”: “%”,
              “value”: “90”
        },
```

```
        {
              “itemCode”: “”,
              “itemName”: “Body_temperature”,
              “measureDateTime”: “”,
              “unit”: “˚C”,
              “value”: “39.4”
        },
        {
              “itemCode”: “”,
              “itemName”: “RR”,
              “measureDateTime”: “”,
              “unit”: “”,
              “value”: “”
        }
        ],
    “symptomList”: [
        {
              “reportDateTime”: “”,
              “text”: “Dry cough, Fatigue, Diarrhea, Slight clinical symptoms, Fever”
        }
    ]
}
```

Discussion

The Template Facilitates the Interoperability in Different Clinical Scenarios

Since the openEHR template developed in our study covered the contents related to clinical characteristics, diagnosis criteria, clinical classification, warning signs for severe and critical cases, differential diagnosis, diagnosis of suspected cases, treatment, and discharge from the latest guideline, it could be used for data exchanging among systems in different clinical scenarios such as screening patients in outpatient clinics, where the diagnosis of suspected cases will be the main focus; the routine round in the wards, where the diagnosis and warning signs for severe or critical cases will be more important; and the intensive care unit, where the treatment recommendation will be the most necessary. Although some hospital information system vendors [8] provided the monolithic solution, the applications for different scenarios with an integrated solution were still encouraged to fully use the expertise of different vendors, which is the norm in the health care institutions. For this reason, the data interoperability is an important issue, and the results of our study can play a significant role.

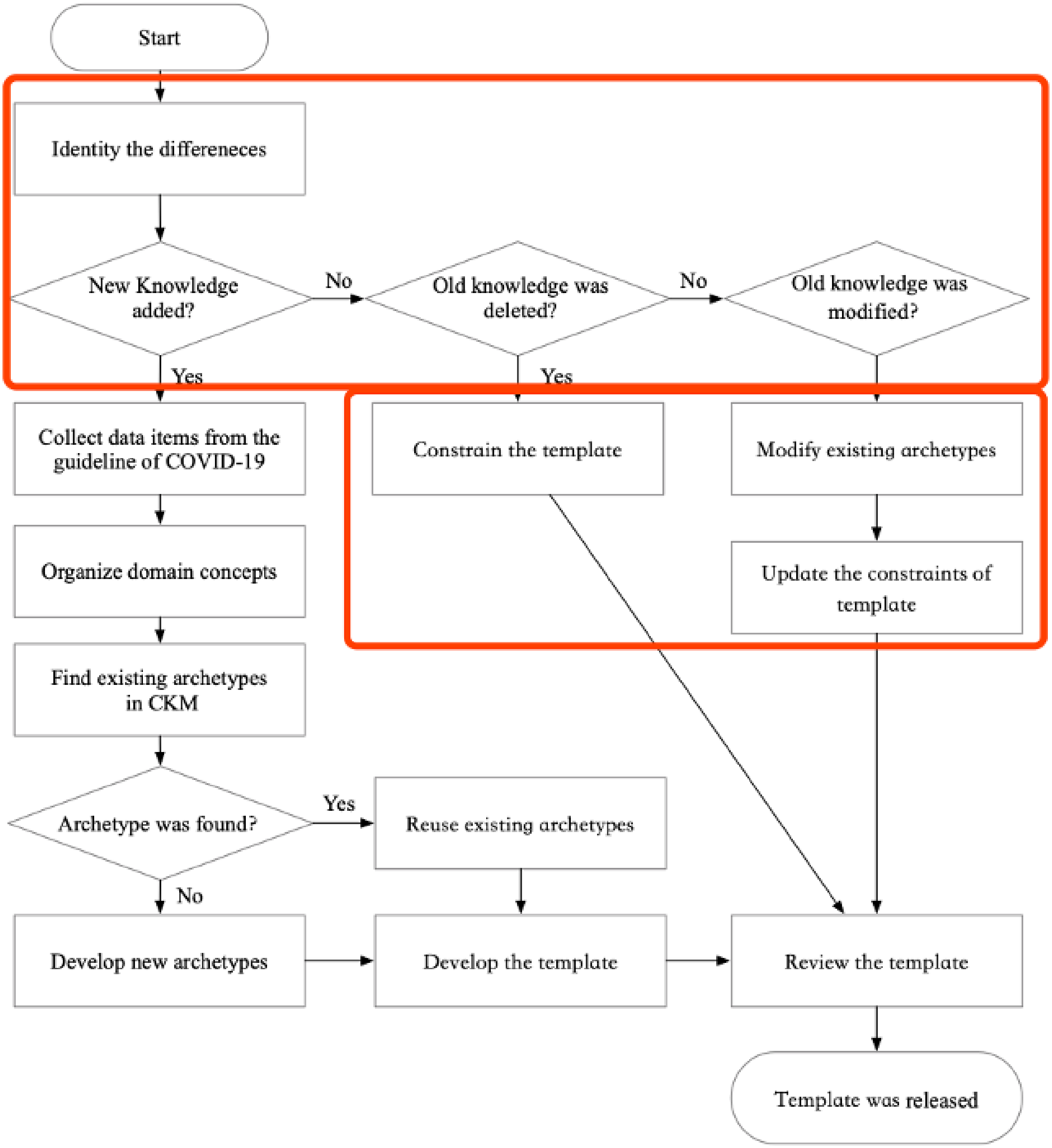
Furthermore, the results of our study can be used for purposes other than the diagnosis and treatment of COVID-19. It can help to develop scales according to severity at different levels and be used for risk assessment of COVID-19. Meanwhile, it is also significant for the prevention and control of the disease in the community. The questionnaires can be designed for people who are under closed management to monitor their physical conditions.

openEHR Modeling Approach Is Flexible for Rapidly Changing Knowledge

COVID-19 was a new threatening infectious disease that brought great pressure on medical systems around the world and with limited previous knowledge in the domain. The methods of diagnosis and treatment have been updated rapidly to reflect the achievement of the latest research since the outbreak of the disease, which sets the challenge for data exchange among systems. The openEHR modeling approach perfectly meets the requirement since the multilevel modeling is especially suitable for the knowledge evolution. In the openEHR ecosystem, when the knowledge of diagnosis and treatment has been updated, only the template needs to be updated and the apps can be kept unchanged. This enables the latest knowledge to be applied to clinical practice at the fastest speed.

In our study, once the latest guideline has been released, the new knowledge can be incorporated into the existing template according to the flowchart shown in Figure 6. Compared with starting from scratch as in the first round, only a few steps need to be performed when the knowledge has been updated, which is shown in the red box (Figure 6).

Figure 6. The updating process of the template. CKM: Clinical Knowledge Manager; COVID-19: coronavirus disease.



The Purpose of Modeling Needs to be Refined for Modeling Precisely

The purpose of the template needs to be refined before modeling since it will largely affect the final results. First, although the purpose of the study is to develop the template for diagnosis and treatment, it still needs to be refined based on whether it is for rule-based decision support tools only or general decision support. As an example, the guideline only mentioned that pregnancy status may affect the intervention without specifying the exact rules, so it is not necessary to be modeled if only for the rule-based decision support tool, but it will still be useful information for professionals to make the decision. Second, the refined purpose should also clarify whether it is used for exchanging the original data from EMR or the condition points for the final decision support. Data items extracted from the guideline usually did not exist in CDR, so they need to be abstracted from the existing data. For instance, the guideline may describe the “Two consecutive negative nucleic acid tests using respiratory tract samples (taken at least 24 hours apart)” as a condition point for discharge, but it has to be calculated from two data items of the nucleic acid test.

Is it Enough to Use the Latest Guideline as the Only Knowledge Source?

This study has used the guideline released by the National Health Commission of the People's Republic of China as the only knowledge source for modeling since the authoritative knowledge was limited at the beginning of the outbreak. However, with the ever-increasing research results and experiences of diagnosis and treatment, there are and will be more knowledge sources, such as the handbook developed by the First Affiliated Hospital, Zhejiang University School of Medicine jointly sponsored by the Jack Ma Foundation and Alibaba Foundation [28], which could be taken as complementary to the guideline. Our modeling approach makes it much easier to apply new or updated knowledge in the template quickly.

More Case Studies and Reviews Aare Needed

Although the template has been reviewed and verified in our study, it still has limitations. First, due to the reason that most of the experienced medical professionals were prioritizing clinical care of patients with COVID 19, it was difficult to have the template reviewed by professionals. Second, since there are not many cases that have been conducted, there may exist some points not appropriate for specific cases (eg, the patient may need to be treated with extra intervention that is not represented in the template). Therefore, further case studies and reviews are necessary to improve the template.

Limitations of Using openEHR Template in an Actual Hospital

The template can be easily deployed in the openEHR-based CDR as shown in this study. However, not all institution's implemented systems are based on openEHR, so there will be a limitation for the use of the template in such scenarios. However, since the infrastructure of openEHR has been designed to be compatible with other existing industry standards, the template can be easily transferred to other popular accepted industry standards like JavaScript Object Notation (JSON), XML, and Health Level 7. To take JSON as an example, the template can be expressed in this format within the support of the JSON schema [29]. The JSON schema is similar to the XML schema to help describe the data format and provide the constraints for the data expression.

Conclusions

This paper developed and released the openEHR template based on the latest guidelines of COVID-19 in China. Most of the archetypes used in the template can be covered by existing archetypes in the CKM. This study proved that the openEHR approach has advantages in modeling a new medical application field and meeting the requirements of rapidly updating knowledge. The template developed in this study could be used to transfer the experience and knowledge achieved from China to other countries and regions as soon as possible from the perspective of improving data exchange among applications to defeat COVID-19.

Acknowledgments

This study was funded by Chinese National Science and Technology Major Project (grant number 2016YFC0901703).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data items extracted from the latest clinical guideline released by China.

[XLSX File (Microsoft Excel File), 15 KB - jmir_v22i6e20239_app1.xlsx]

Multimedia Appendix 2

The details of domain concepts presented in mind map.

[PDF File (Adobe PDF File), 224 KB - jmir_v22i6e20239_app2.pdf]

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission Dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
2. National Health Commission of the People's Republic of China. Class B infectious diseases were managed according to the Class A infectious diseases URL: http://www.nhc.gov.cn/xcs/pfzs/202002/b4813a3a223c475c993f957b30f298b9.shtml [accessed 2020-03-18]
3. Chinese Center for Disease Control and Prevention. The classification of infectious diseases in China URL: http://www.chinacdc.cn/jkzt/crb/index.html [accessed 2020-03-18]
4. World Health Organization. 2020 Jan 23. Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) URL: https://tinyurl.com/v8ulkhf [accessed 2020-03-01]
5. World Health Organization. Coronavirus disease (COVID-2019) situation reports URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [accessed 2020-03-04]

6. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
7. World Health Organization. Global research on coronavirus disease (COVID-19) URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov [accessed 2020-03-18]
8. Miliard M. Healthcare IT News. 2020 Jan 24. Epic pushes out software update to help spot coronavirus URL: https://www.healthcareitnews.com/news/epic-pushes-out-software-update-help-spot-coronavirus [accessed 2020-03-24]
9. openEHR. 2020 Mar 11. openEHR community rises to the challenge of coronavirus URL: https://www.openehr.org/news_events/openehr_news/311 [accessed 2020-03-24]
10. SNOmed International. 2020 Mar. March 2020 SNOMED CT International Edition Interim Release: up to date COVID-19 content available URL: http://www.snomed.org/news-and-events/articles/march-2020-interim-snomedct-release-COVID-19 [accessed 2020-03-24]
11. GitHub. The OHDSI repository to provide comprehensive evidence for the covid-19 URL: https://github.com/OHDSI/Covid-19 [accessed 2020-03-24]
12. Schloeffel P, Beale T, Hayworth G, Heard S, Leslie H. The relationship between CEN 13606, HL7, and openEHR. HIC 2006 and HINZ 2006: Proceedings, 24 2006:1-4.
13. Goossen W, Goossen-Baremans A, van der Zel M. Detailed clinical models: a review. Healthc Inform Res 2010 Dec;16(4):201-214 [FREE Full text] [doi: 10.4258/hir.2010.16.4.201] [Medline: 21818440]
14. Jiang G, Xiao G, Kiefer RC, Prod'hommeaux E, Solbrig HR. Building an FHIR ontology based data access framework with the OHDSI data repositories. AIMA 2017:1-2.
15. Jiang G, Prud'Hommeaux E, Xiao G, Solbrig HR. Developing a semantic web-based framework for executing the clinical quality language using FHIR. CEUR-WS.org 2017:1-5.
16. openEHR Foundation. openEHR architecture overview URL: https://www.openehr.org/releases/BASE/latest/docs/architecture_overview/architecture_overview.html [accessed 2020-03-01]
17. Thurston LM. Flexible and extensible display of archetyped data: the openEHR presentation challenge. HIC 2006 and HINZ 2006: Proceedings 2006:28-36.
18. National Health Commission of the People's Republic of China. Novel coronavirus pneumonia diagnosis and treatment plan (provisional 7th edition) URL: http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml [accessed 2020-03-24]
19. China Law Translate. 2020 Mar 04. Novel coronavirus pneumonia diagnosis and treatment plan (provisional 7th edition) URL: https://www.chinalawtranslate.com/en/coronavirus-treatment-plan-7/.[accessed [accessed 2020-03-24]
20. openEHR. Clinical Knowledge Manager URL: https://www.openehr.org/ckm/ [accessed 2020-03-18]
21. openEHR. Archetype definition language URL: https://specifications.openehr.org/releases/AM/latest/ADL1.4.html [accessed 2020-03-18]
22. openEHR. Ocean Template Editor URL: https://www.openehr.org/downloads/modellingtools/ [accessed 2020-03-18]
23. Zhejiang University. Healthcare Modeling Collaboration URL: http://openehr.org.cn/hmc/ [accessed 2020-03-01]
24. Min L, Liu J, Lu X, Duan H, Qiao Q. An implementation of Clinical Data Repository with openEHR approach: from data modeling to architecture. Stud Health Technol Inform 2016;227:100-105. [Medline: 27440296]
25. Wang L, Min L, Wang R, Lu X, Duan H. Archetype relational mapping - a practical openEHR persistence solution. BMC Med Inform Decis Mak 2015 Nov 05;15:88 [FREE Full text] [doi: 10.1186/s12911-015-0212-0] [Medline: 26541142]
26. Frade S, Freire SM, Sundvall E, Patriarca-Almeida JH, Cruz-Correia R. Survey of openEHR storage implementations. 2013 Presented at: 26th IEEE International Symposium on Computer-Based Medical Systems; 20-22 June 2013; Porto, Portugal. [doi: 10.1109/cbms.2013.6627806]
27. openEHR. COVID-19 pneumonia diagnosis and treatment (7th edition) URL: https://ckm.openehr.org/ckm/templates/1013.26.291
28. Handbook of COVID-19 Preventionand Treatment. URL: https://www.alibabacloud.com/universal-service/pdf_reader?spm=a3c0i.14138300.8102420620.dreadnow.6df3647fijaJIV&cdnorigin=video-intl&pdf=Read%20Online-Handbook%20of%20COVID-19%20Prevention%20and%20Treatment.pdf [accessed 2020-03-24]
29. JSON Schema. URL: https://json-schema.org/ [accessed 2020-03-24]

Abbreviations

API: application programming interface
CDR: clinical data repository
CDSS: clinical decision support system
CKM: Clinical Knowledge Manager
COVID-19: coronavirus disease
JSON: JavaScript Object Notation

REST: representational state transfer
RM: reference model
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.05.20; peer-reviewed by S Kobayashi, I McNicoll, N Deng; comments to author 27.05.20; revised version received 03.06.20; accepted 03.06.20; published 10.06.20.

Please cite as:
Li M, Leslie H, Qi B, Nan S, Feng H, Cai H, Lu X, Duan H
Development of an openEHR Template for COVID-19 Based on Clinical Guidelines
J Med Internet Res 2020;22(6):e20239
URL: http://www.jmir.org/2020/6/e20239/
doi:10.2196/20239
PMID:32496207

Original Paper

Epidemiological Observations on the Association Between Anosmia and COVID-19 Infection: Analysis of Data From a Self-Assessment Web Application

Fabrice Denis[1], MD, PhD; Simon Galmiche[2], MD; Aurélien Dinh[3], MD; Arnaud Fontanet[2], MD, PhD; Arnaud Scherpereel[4], MD, PhD; Francois Benezit[5], MD; François-Xavier Lescure[6,7], MD, PhD

[1]Inter-regional Cancer Institut Jean Bernard, Le Mans, France

[2]Emerging Diseases Epidemiology Unit, Institut Pasteur, Paris, France

[3]Service de maladies infectieuses et tropicales, Hôpital Raymond Poincaré, Assistance Publique - Hôpitaux de Paris, Garches, France

[4]Service de pneumologie, Centre Hospitalier Régional Universitaire de Lille, Lille, France

[5]Service de maladies infectieuses et réanimation médicale, Centre Hospitalier Universitaire de Rennes Pointchaillou, Rennes, France

[6]Infectious and Tropical Diseases Department, Bichat-Claude Bernard University Hospital and University of Paris, Assistance Publique - Hôpitaux de Paris, Paris, France

[7]Team DesCID, Infection, Antimicrobials, Modelling, Evolution - U1137, French Institute for Health and Medical Research, Institut national de la santé et de la recherche médicale, Paris, France

Corresponding Author:
Fabrice Denis, MD, PhD
Inter-regional Cancer Institut Jean Bernard
9, rue Beauverger
Le Mans, 72100
France
Phone: 33 243475810
Email: f.denis@cjb72.org

Abstract

Background: We developed a self-assessment and participatory surveillance web application for coronavirus disease (COVID-19), which was launched in France in March 2020.

Objective: Our objective was to determine if self-reported symptoms could help monitor the dynamics of the COVID-19 outbreak in France.

Methods: Users were asked questions about underlying conditions, sociodemographic status, zip code, and COVID-19 symptoms. Depending on the symptoms reported and the presence of coexisting disorders, users were told to either stay at home, contact a general practitioner (GP), or call an emergency phone number. Data regarding COVID-19–related hospitalizations were retrieved from the Ministry of Health.

Results: As of March 29, 2020, the application was opened 4,126,789 times; 3,799,535 electronic questionnaires were filled out; and 2,477,174 users had at least one symptom. In total, 34.8% (n=1,322,361) reported no symptoms. The remaining users were directed to self-monitoring (n=858,878, 22.6%), GP visit or teleconsultation (n=1,033,922, 27.2%), or an emergency phone call (n=584,374, 15.4%). Emergency warning signs were reported by 39.1% of participants with anosmia, a loss of the sense of smell (n=127,586) versus 22.7% of participants without anosmia (n=1,597,289). Anosmia and fever and/or cough were correlated with hospitalizations for COVID-19 (Spearman correlation coefficients=0.87 and 0.82, respectively; *P*<.001 for both).

Conclusions: This study suggests that anosmia may be strongly associated with COVID-19 and its severity. Despite a lack of medical assessment and virological confirmation, self-checking application data could be a relevant tool to monitor outbreak trends.

Trial Registration: ClinicalTrials.gov NCT04331171; https://clinicaltrials.gov/ct2/show/NCT04331171

(J Med Internet Res 2020;22(6):e19855) doi:10.2196/19855

KEYWORDS

COVID-19; anosmia; epidemiological surveillance; self-assessment; web application; outbreak; symptoms; self-assessment; surveillance; epidemiology

Introduction

Web-based self-reporting of symptoms is a growing field and has been used to improve survival in oncology [1,2]; it can be used as a participatory surveillance tool for coronavirus disease (COVID-19) or other influenza-like illnesses as well [3,4]. We thought of applying the same technology to optimize patient triage for COVID-19 patients in France and alleviate the burden on emergency call centers. A self-assessment and participatory surveillance website [5] was developed and launched during the growing phase of the COVID-19 epidemic in France in March 2020. Our objective was to determine if self-reported symptoms could help monitor outbreak dynamics in France. We report here the analysis of the first 13 days of web application usage.

Methods

Users were recruited via national media campaigns in France, including social media, radio, and magazine campaigns, between March 17-29, 2020. Participants were recruited through the maladiecoronavirus.fr website [5]. Respondents provided information on sociodemographic data, zip code, and coexisting disorders anonymously. They were asked about nine symptoms associated with possible COVID-19 infection—fever (body temperature >37.7°C), unusual cough, shortness of breath, sore throat, muscle aches, diarrhea, anorexia, and asthenia. Anosmia, a loss of the sense of smell, was added on March 21, 2020. Following symptom reporting, a notification was sent, recommending the user either to stay at home and use the website again in case of evolving symptomatology (self-monitoring), or to contact a general practitioner (GP), or to call an emergency number if they reported dyspnea or anorexia. A questionnaire was built according to Chinese reports and French experience [6]. The website was not considered a medical device by regulatory authorities since no tracking was performed and data were anonymous. We compared the distribution of web-based, self-reported symptoms to that of hospitalized COVID-19 patients according to Ministry of Health reports. Spearman correlation coefficients were used for statistical analysis.

Results

Between March 17-29, 2020, the website was accessed 4,126,789 times; 3,799,535 electronic questionnaires were filled out; and 2,477,174 users had at least one out of the nine symptoms included in the questionnaire (Figure 1). In total, 1,322,361 (34.8%) participants reported no symptoms. The remaining patients (median age 37 years; range 15-99 years) were directed to self-monitoring (858,878, 22.6%), GP visit or teleconsultation (1,033,922, 27.2%), or an emergency phone call (584,374, 15.4%).

Of all symptomatic patients, anosmia was reported by 17.1% (325,910/1,903,741), fever was reported by 33.5% (828,952/2,477,174), and cough by 61.2% (1,515,557/2,477,174). Emergency warning signs (dyspnea or complete anorexia) were reported by 39.1% of participants with anosmia (n=127,586) versus 22.7% of participants without anosmia (n=1,597,289; *P*<.001; Table 1). Anosmia and fever and/or cough were correlated with COVID-19–related hospitalizations (Spearman correlation coefficients=0.87 and 0.82, respectively; *P*<.001 for both; Figure 2).

Figure 1. Flowchart of the study population.

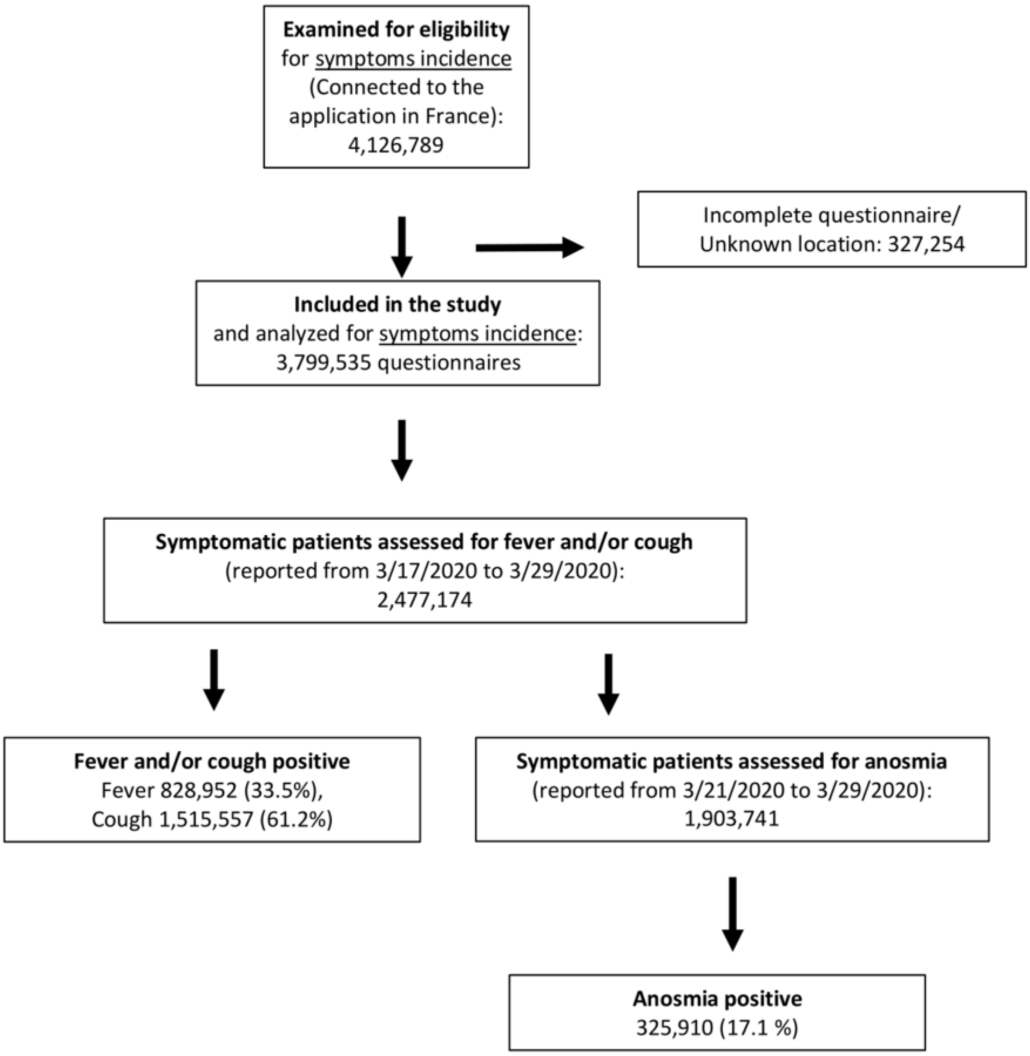


Table 1. Users characteristics.

Characteristic	Value (n=2,477,174[a])
Age (years), average (range), median	39.12 (15-99), 37
Sex	—[b]
Body mass index (kg/m^2)	
≥30	436,609 (17.6)
<30	2,040,555 (82.4)
Comorbidities	
Cardiovascular disease / uncontrolled hypertension	401,888 (16.2)
Diabetes	78,354 (3.2)
Malignancy	72,577 (2.9)
Pulmonary disease	243,247 (9.8)
Chronic kidney disease	11,333 (0.5)
Chronic liver disease	36,677 (1.5)
Pregnancy	40,766 (1.6)
Immunodepression	135,624 (5.5)
Immunosuppressive ongoing therapy	70,927 (2.9)
Reported symptoms	
Fever (body temperature greater >37.7°C)	828,952 (33.5)
Cough	1,515,557 (61.2)
Dyspnea	658,442 (26.6)
Asthenia	1,155,297 (46.6)
Complete anorexia	103,122 (4.2)
Sore throat or muscle aches	1,837,286 (74.1)
Diarrhea	497,665 (20.1)
Anosmia or dysgeusia[c]	325,910 (17.1)
Patient triage after questionnaire completion[d]	
No symptoms	1,322,361 (34.8)
Self-monitoring	858,878 (22.6)
General practitioner / phone call	1,033,922 (27.2)

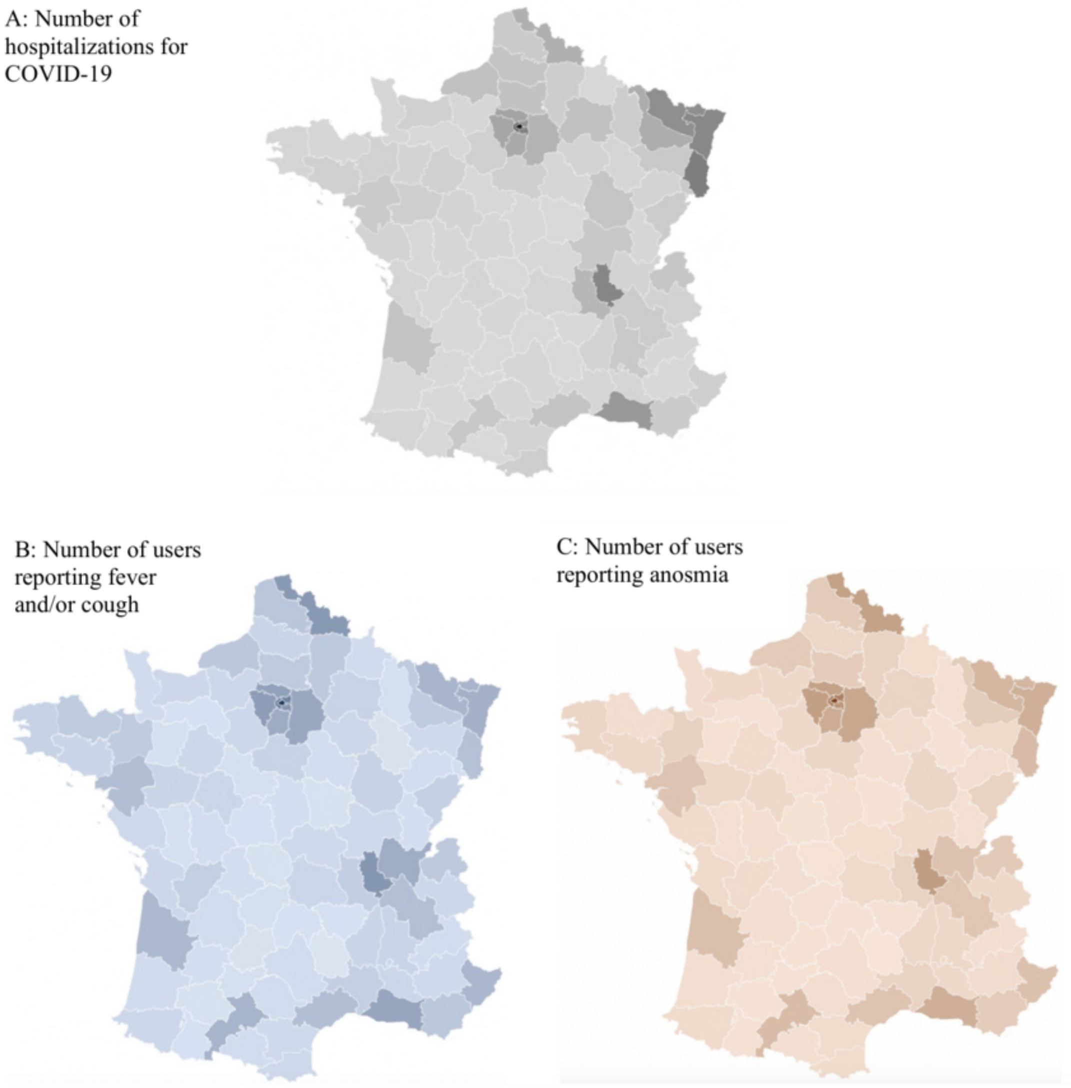
[a] Symptomatic patient (reported at least one symptom).

[b] Sex was not asked to protect the identity of the user.

[c] Anosmia was reported from March 21, 2020; the number of symptomatic patients during this period was 1,903,741.

[d] Denominator is N=3,799,535 (all respondents; symptomatic or not).

Figure 2. Maps displaying the correlation between fever and/or cough and anosmia with hospitalizations for COVID-19. (A) The cumulative number of hospitalized COVID-19–positive patients in France on March 29, 2020. (B) Fever and/or cough reported by users via the application; cumulative amount in French counties from March 17-29, 2020: 3,799,535 respondents (828,952 with fever and 1,515,557 with cough). (C) Anosmia reported by users; cumulative number from March 21-29, 2020: 325,910 positive respondents.



A: Number of hospitalizations for COVID-19

B: Number of users reporting fever and/or cough

C: Number of users reporting anosmia

Discussion

This study suggests that self-reported symptoms of COVID-19 are correlated with COVID-19–related hospitalizations and that anosmia may be strongly associated with COVID-19. This could be explained by a greater specificity compared to other reported symptoms that could result from other respiratory viruses [7,8].

Limitations include lack of medical assessment and virological confirmation of COVID-19 and comparison of application-retrieved data with a distinct set of data from the Ministry of Health reports on hospitalizations. Self-checking application data could be a relevant tool to monitor the dynamics of an outbreak and thus can be a real-time health system response to the epidemic.

Conflicts of Interest

FD received personal fees from AstraZeneca, Ipsen, Sivan Innovation, Pfizer, Chugai, Takeda, and Roche. All other authors declared no conflicts of interest.

References

1. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]

2. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
3. Luo H, Lie Y, Prinzen FW. Surveillance of COVID-19 in the General Population Using an Online Questionnaire: Report From 18,161 Respondents in China. JMIR Public Health Surveill 2020 Apr 27;6(2):e18576 [FREE Full text] [doi: 10.2196/18576] [Medline: 32319956]
4. Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory Syndromic Surveillance of Influenza in Europe. J Infect Dis 2016 Dec 01;214(suppl_4):S386-S392. [doi: 10.1093/infdis/jiw280] [Medline: 28830105]
5. maladiecoronavirus.fr. 2020. URL: https://maladiecoronavirus.fr/ [accessed 2020-03-17]
6. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
7. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 May 11. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
8. Bénézit F, Le Turnier P, Declerck C, Paillé C, Revest M, Dubée V, et al. Utility of hyposmia and hypogeusia for the diagnosis of COVID-19. The Lancet Infectious Diseases 2020 Apr 15 [FREE Full text] [doi: 10.1016/S1473-3099(20)30297-8] [Medline: 32304632]

Abbreviations

COVID-19: coronavirus disease
GP: general practitioner

Edited by G Eysenbach; submitted 04.05.20; peer-reviewed by F Lanternier, E Stavaux, O Leal Neto; comments to author 27.05.20; revised version received 02.06.20; accepted 03.06.20; published 11.06.20.

Please cite as:
Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, Lescure FX
Epidemiological Observations on the Association Between Anosmia and COVID-19 Infection: Analysis of Data From a Self-Assessment Web Application
J Med Internet Res 2020;22(6):e19855
URL: http://www.jmir.org/2020/6/e19855/
doi:10.2196/19855
PMID:32496206

Original Paper

Information Technology–Based Management of Clinically Healthy COVID-19 Patients: Lessons From a Living and Treatment Support Center Operated by Seoul National University Hospital

Ye Seul Bae[1,2], MD, MPH; Kyung Hwan Kim[1,3,4], MD, PhD; Sae Won Choi[1,5], MD; Taehoon Ko[1], PhD; Chang Wook Jeong[1,6], MD, PhD; BeLong Cho[2,7], MD, MPH, PhD; Min Sun Kim[7,8], MD; EunKyo Kang[2,7], MD

[1]Office of Hospital Information, Seoul National University Hospital, Seoul, Republic of Korea
[2]Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[3]Department of Thoracic and Cardiovascular Surgery, Seoul National University Hospital, Seoul, Republic of Korea
[4]Department of Thoracic and Cardiovascular Surgery, Seoul National University College of Medicine, Seoul, Republic of Korea
[5]Department of Emergency Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[6]Department of Urology, Seoul National University Hospital, Seoul, Republic of Korea
[7]Department of Public Health and Medical Service, Seoul National University Hospital, Seoul, Republic of Korea
[8]Department of Pediatrics, Seoul National University Hospital, Seoul, Republic of Korea

Corresponding Author:
Kyung Hwan Kim, MD, PhD
Office of Hospital Information
Seoul National University Hospital
101 Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 2 2072 7600
Fax: 82 2 2072 7600
Email: kkh726@snu.ac.kr

Abstract

Background: South Korea took preemptive action against coronavirus disease (COVID-19) by implementing extensive testing, thorough epidemiological investigation, strict social distancing, and rapid treatment of patients according to disease severity. The Korean government entrusted large-scale hospitals with the operation of living and treatment support centers (LTSCs) for the management for clinically healthy COVID-19 patients.

Objective: The aim of this paper is to introduce our experience implementing information and communications technology (ICT)-based remote patient management systems at a COVID-19 LTSC.

Methods: We adopted new electronic health record templates, hospital information system (HIS) dashboards, cloud-based medical image sharing, a mobile app, and smart vital sign monitoring devices.

Results: Enhancements were made to the HIS to assist in the workflow and care of patients in the LTSC. A dashboard was created for the medical staff to view the vital signs and symptoms of all patients. Patients used a mobile app to consult with their physician or nurse, answer questionnaires, and input self-measured vital signs; the results were uploaded to the hospital information system in real time. Cloud-based image sharing enabled interoperability between medical institutions. Korea's strategy of aggressive mitigation has "flattened the curve" of the rate of infection. A multidisciplinary approach was integral to develop systems supporting patient care and management at the living and treatment support center as quickly as possible.

Conclusions: Faced with a novel infectious disease, we describe the implementation and experience of applying an ICT-based patient management system in the LTSC affiliated with Seoul National University Hospital. ICT-based tools and applications are increasingly important in health care, and we hope that our experience will provide insight into future technology-based infectious disease responses.

(J Med Internet Res 2020;22(6):e19938) doi:10.2196/19938

KEYWORDS

COVID-19; clinical informatics; mobile app; telemedicine; hospital information system; app; health information technology

Introduction

The coronavirus disease (COVID-19) pandemic has become a major concern worldwide since the first report from Wuhan, China in December 2019 [1-4]. According to the World Health Organization, up to 4 million patients and 277,000 deaths have been reported [5,6]. In South Korea, the first confirmed case of COVID-19 was diagnosed on January 20, 2020; as of May 10, 2020, 10,874 patients have been diagnosed, 9610 have been released from isolation, and 256 have died [7].

South Korea has earned recognition worldwide for its tackling of the pandemic. The number of infections dropped from 900 a day in late February to less than 20 per day in late April. South Korea acted preemptively against COVID-19 with robust testing, vigorous tracing, strict social distancing, and rapid treatment of patients according to severity. In late February, a super-spreader, "Patient 31," caused a regional outbreak in the southern city of Daegu and Gyeongsangbuk Province by attending religious services. Health authorities conducted contact tracing and decided to test every close contact irrespective of whether they showed any symptoms [8]. Up to 80% of patients tested were clinically healthy, and due to the substantial number of tests performed, medical institutions were highly saturated. To provide appropriate treatment and to reduce the burden on the medical institutions of Daegu and Gyeongsangbuk Province, the government established a new treatment system based on severity on March 1, 2020 [9]. Every patient was classified based on severity into mild, moderate, severe, or extremely severe cases. Patients identified as moderate, severe, and extremely severe were immediately hospitalized for treatment [9,10].

Asymptomatic patients testing positive for COVID-19 and mild symptomatic patients were isolated and accommodated at government-sponsored facilities called living and treatment support centers (LTSCs) to be monitored by health care staff at least twice a day [9]. Patients can be discharged from the LTSC after two consecutive negative tests. In case of emergencies, patients are transferred to nearby hospitals [9]. The government asked several tertiary hospitals to operate LTSCs. The first LTSC opened in Daegu; Seoul National University Hospital (SNUH) opened the third LTSC at the SNUH Human Resource Development Center in Mungyeong, Gyeongsangbuk Province, 180 kilometers southeast of Seoul and 100 kilometers northwest of Daegu.

In this study, we introduce the experience of operating a LTSC for the management of mild COVID-19 patients and introducing information and communications technology (ICT)–based solutions tailored to the characteristics of the patients' clinical pathways.

Methods

We implemented new electronic health record (EHR) templates for the hospital information system (HIS), dashboards, an electronic prescription (e-prescription) system, and cloud-based medical image sharing. A newly developed mobile app enabled effective and accurate communication between health care providers and patients, and wearable vital sign monitoring devices facilitated management of clinically healthy COVID-19 patients.

The SNUH HIS

SNUH is a non-profit, academic referral government hospital with 1778 beds. SNUH treats approximately 600,000 inpatients and 2.3 million outpatients per year and employs about 1400 physicians and 2200 nurses. SNUH internally developed a computerized physician order entry system in 1999, implemented a picture archiving and communication system (PACS) in 2001, and started using an EMR in 2004. In 2001, SNUH's internal information technology (IT) department was spun off as a separate company, ezCaretech, which is now the largest HIS company in Korea [11]. The current version of the HIS, BESTCare 2.0, which integrates and manages all services needed by the hospital, was built by ezCaretech and implemented in 2016. The HIS was modified to support the LTSC.

Patient Management Systems to Support a LTSC in Mungyeong

Health authorities screened and designated clinically healthy COVID-19 patients to be admitted to LTSCs; however, patient management was entirely the responsibility of the entrusted hospital. Patients were considered clinically healthy if their vital signs were stable (blood pressure, heart rate, and oxygen saturation within normal limits), if they were afebrile, and if they were asymptomatic or had mild COVID-19 symptoms. Clinical health status was determined by directly evaluating the patient at the time of diagnosis. Blood pressure monitors, pulse oximeters, and thermometers were placed in each room for the patient to measure their blood pressure, oxygen saturation, and temperature. In some rooms, a wearable continuous vital sign monitoring device was placed alongside conventional devices for continuous remote monitoring. To protect the medical staff, exposure was minimized via the use of remote communication methods and self-monitoring by the patients.

User Survey

We conducted online surveys to measure satisfaction with the systems modified for the LTSCs. Three separate online surveys were conducted: patient satisfaction with the mobile app, patient satisfaction with the wearable device, and satisfaction of the medical staff. Only participants who voluntarily agreed to participate were included. Responses were measured using a 5-point Likert scale. The questionnaires were administered using an online survey form. Participants accessed questionnaires via a URL and were able to complete the survey at any time or place. The study was approved by the Institutional Review Board of SNUH (IRB number: H-2004-026-1115). The ethics committee waived the need for informed participant consent.

Results

Patient Management in the LTSC

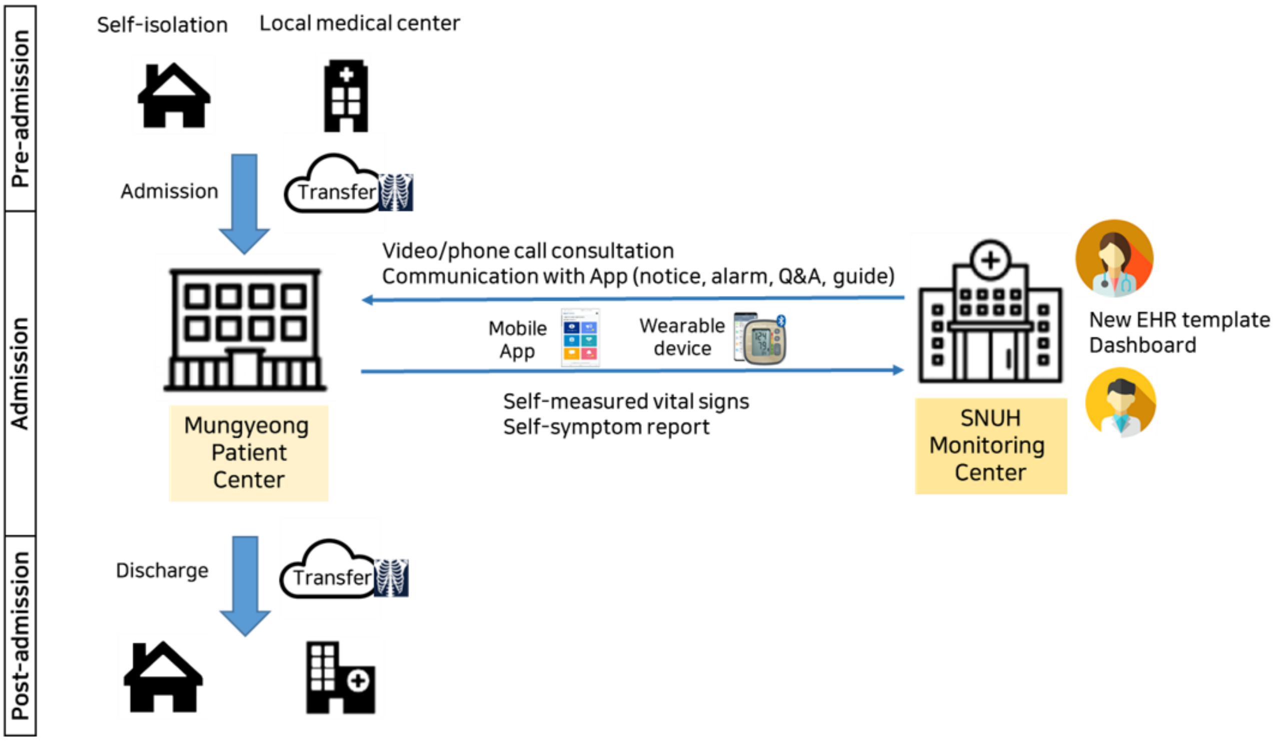
Patients who were positive for the COVID-19 real time polymerase–chain reaction (RT-PCR) test but were asymptomatic or had mild symptoms were subject to LTSC admission. Upon disease confirmation, a group of experts from public authorities triaged the patient and determined treatment depending on the severity of the patient's symptoms. Patients with severe symptoms were hospitalized and treated in negative pressure isolation units, and mild and asymptomatic patients were designated to be admitted to LTSCs [7-9]. Patients were initially admitted to the LTSC from self-isolation in their homes or were transferred from other local hospitals.

As of April 20, 2020, a total of 18 LTSCs were operational in Korea. 2948 patients were admitted to the centers, of whom 155 were transferred to another hospital due to worsening symptoms [12]. Because the number of newly diagnosed patients decreased and the number of people who were released from quarantine increased, the 18 LTSCs were consolidated into 5 centers in May 2020.

The SNUH facility in Mungyeong was originally constructed as a human resource development center used to train SNUH personnel. It was not a medical facility and therefore did not contain any medical equipment. After the decision was made to adapt the facility at Mungyeong into a LTSC, the necessary equipment, including personal computers, monitors, a portable chest x-ray machine, medical devices, and the hospital network, was installed in 2 days. Since the LTSC opened on March 5, 2020, a total of 118 patients have been admitted.

Figure 1 shows the overall flow of ICT operations according to the patient journey. A cloud system was used for instant medical image sharing in cases of patient transfer. Patients self-reported vital signs and subjective symptoms through a mobile app. These data were automatically linked to the newly introduced semistructured EHR template for mild COVID-19 patients in the HIS to help the medical staff obtain patient-reported data conveniently and accurately. In addition, in some rooms equipped with wearable devices, vital signs were transmitted to the HIS in real time. Through the mobile app, medical staff were able to notify and alert individual patients and communicate with them when needed. The Korean government temporarily permitted virtual patient management to minimize COVID-19 transmission on February 22, 2020 [7]. Each patient self-measured vital signs and reported symptoms through the mobile app at 9:00 AM and 4:30 PM every day, and they spoke with the nurse in charge via remote video consult twice a day from 9 AM-12 PM and 5 PM-8 PM and once a day with a physician in Seoul. If the patient had a concern or problem, the assigned nurse could be contacted by text or telephone. To respond to emergencies, more than two medical staff members (physicians or nurses) were always present in the patient center at Mungyeong.

Figure 1. Overall ICT operation flow according to the patient's journey. EHR, electronic health record. ICT: information and communications technology. Q&A: question and answer. SNUH: Seoul National University Hospital.



Patients were discharged if they had two consecutive negative COVID-19 tests. The RT-PCR test used by SNUH is the STANDARD M nCov Real-Time Detection Kit (SD Biosensor), which was issued an emergency use authorization by the United States Food and Drug Administration and the Korea Food and Drug Administration. The sensitivity of the test is 100% and the specificity is greater than 97% [13,14]. Chest x-rays were performed periodically to check for progression to pneumonia after admission [7-9].

The Korean government funded the operation of the LTSCs through reimbursement of medical fees. The conditions for reimbursement were as follows: medical personnel are dispatched to the LTSC, patients are seen twice a day, and necessary medical equipment, such as an x-ray machine, is available. The fee of about US $30 per patient per day could be claimed if the conditions were met [15].

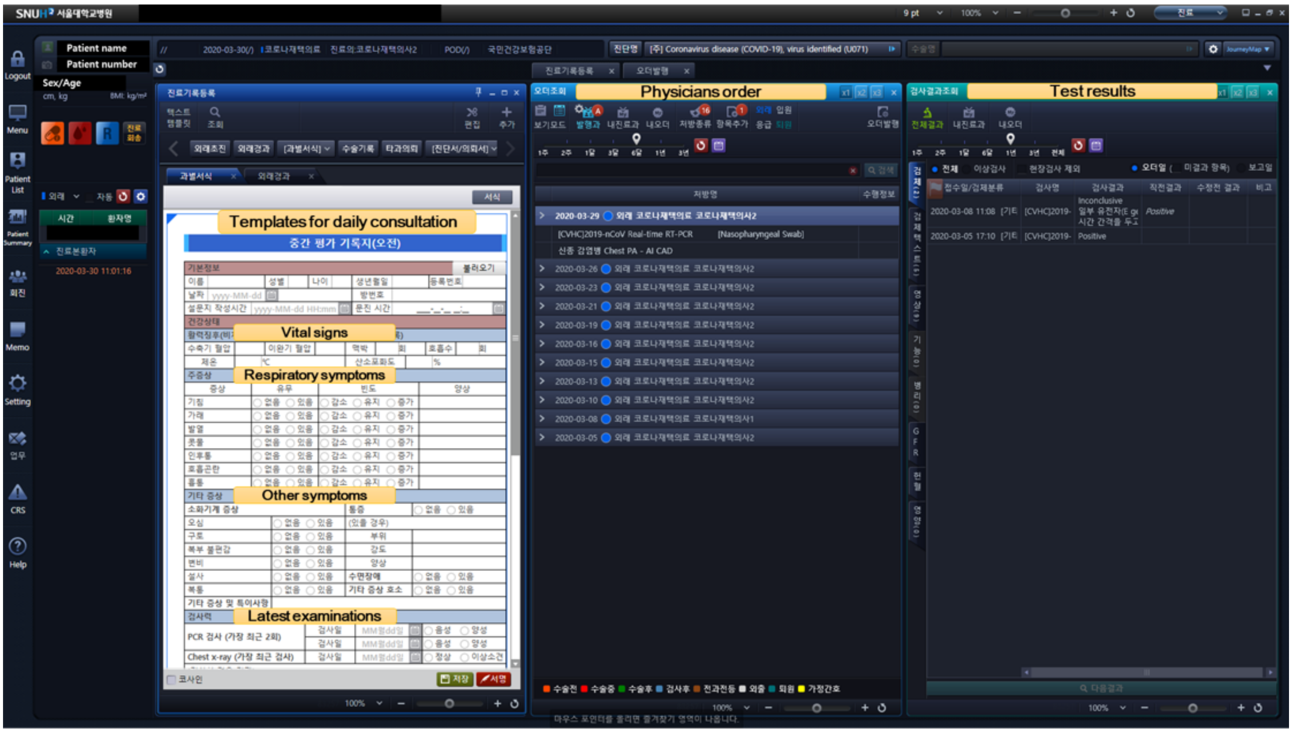
Enhancements to the HIS

A multidisciplinary task force of physicians, nurses, infection specialists, epidemiologists, and IT experts headed by the chief information officer of the hospital was convened to direct and manage the modifications to the IT systems to support the LTSC. First, we customized and enhanced our HIS to support the LTSC.

New EHR Templates and Order Codes

To use the SNUH HIS in the LTSC, a hospital network was installed in addition to the usual internet service. An LTSC department was established in the HIS for efficient and independent patient management. Only chest x-rays and RT-PCR using a nasopharyngeal swab or sputum samples were performed at the LTSC; new order codes were created for these tests. Moreover, we developed three new EHR templates for admission, progression, and discharge, which were used to track the symptoms and general status of the patients (Figure 2). The EHR templates for the LTSCs were constructed in a semistructured form that combined a free text section for medical staff to freely record observations during consultation and a structured form that was configured to include patients' subjective respiratory, gastrointestinal, and psychological symptoms, objective vital signs, previous medical history, contact history of COVID-19, and allergies. When a patient entered self-reported symptoms or vital signs on the mobile app, the data were automatically linked to the EHR template to be viewed during remote video consultation.

Figure 2. Modified EHR template for daily consultation at the LTSC.



Dashboard

A dashboard for the medical staff was developed to display real time information for all patients at central monitoring centers at Seoul and Mungyeong. The dashboard displayed each patient's vital signs (heart rate, body temperature, respiratory rate, blood oxygen saturation) and information on whether the patient has any symptoms or chest x-ray abnormalities (Figure 3).

Figure 3. Dashboard of the EHR displaying real time information for patients at the LTSC.

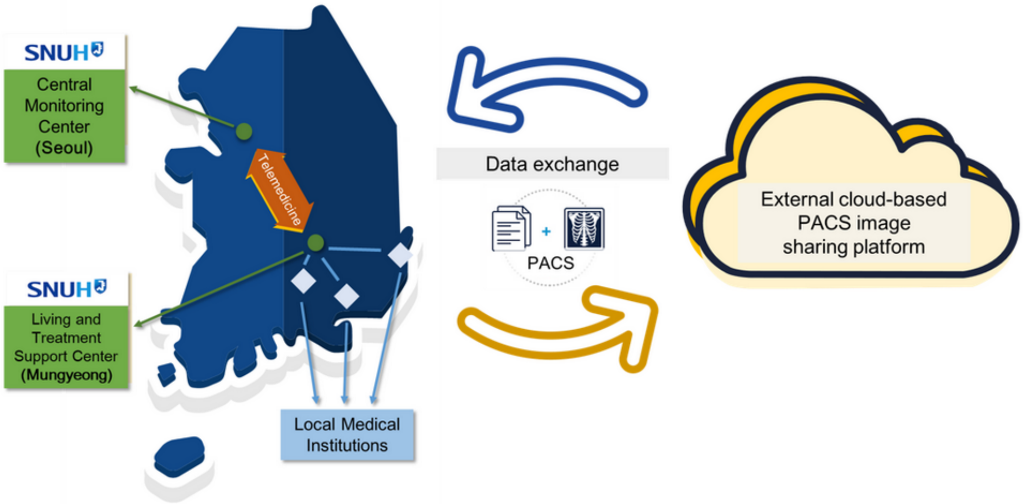
e-Prescription System

Basic medications were prepared at SNUH and transported to the LTSC facility. Plans were made for sourcing other medications from local pharmacies by sending prescriptions electronically. A mobile app–based e-prescription system was added to the SNUH app for patients on March 1, 2020. The patient could send the prescription information through the app to a nearby pharmacy selected by the user. There were two cases of e-prescription system usage in the LTSC. A 25-year-old female patient taking propranolol regularly for palpitation was admitted but did not bring the medication with her. During the stay, she complained of palpitations with tachycardia. The physician at the central monitoring center in Seoul prescribed propranolol, and the prescription was electronically sent through the app to a local pharmacy about 1.8 km from the LTSC. Another case was a 25-year-old male patient with hyperthyroidism who was prescribed and received methimazole using the e-prescription system.

Cloud-Based Medical Image Sharing

Seamless medical image interoperability is not available in Korea. Generally, medical images are copied to CDs and carried by the patient from one institution to another. Clearly, this process is inadvisable due to the nature of the disease. In addition, SNUH decided to review the chest x-ray and computerized tomography images taken by radiologists at SNUH before a to the LTSC. Also, producing a CD copy for patients in need of transfer to another hospital due to emergent or worsening conditions was not practical. Thus, we decided to establish a simple cloud-based system for instant and effective medical image sharing (Figure 4). Users from other hospitals would upload the patient's medical images to the cloud, which were then transferred to the SNUH PACS for physician review. All medical images were shared using the Digital Imaging and Communications in Medicine (DICOM) format. Twelve patients used this system when they were admitted. Two patients who had no symptoms at the time of admission developed difficulty breathing and were transferred to a nearby hospital. Chest x-rays taken at LTSC were quickly shared with medical staff in the receiving hospital using this system.

Figure 4. The cloud-based medical image sharing platform. PACS: picture archiving and communication system. SNUH: Seoul National University Hospital.



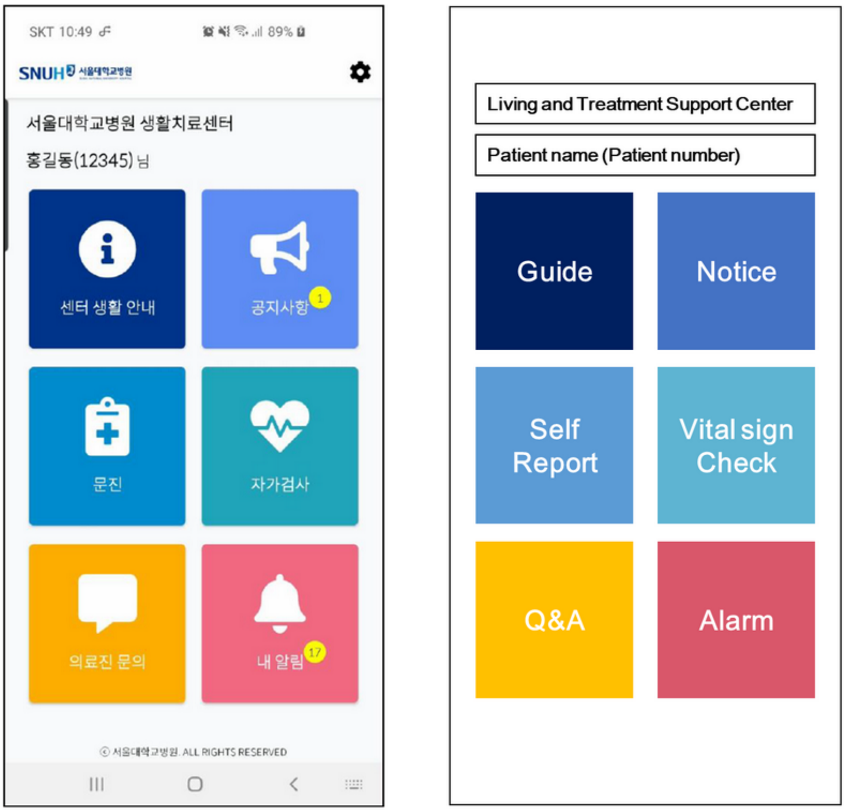
Wearable Vital Sign Monitoring Device

Some patients were asked to use a wearable device to measure their vital signs and enable the medical staff to monitor them remotely. The Vital-sign Data Recorder (VDR-1000) from Tribell Labs is a wearable medical multi-function measurement device that can be easily used by patients (Multimedia Appendix 1). It can directly measure the patient's electrocardiogram, pulse rate, blood oxygen saturation, respiratory waveform, respiratory rate, and body temperature. The measured data are transmitted to the central monitoring system (CMS) using Wi-Fi and are then forwarded to the HIS. Medical staff in Mungyeong and Seoul can monitor vital signs of patients using both the CMS monitor and HIS (Multimedia Appendix 2). The CMS software can set alarms with different thresholds for each patient. An alarm sounds when a value outside the threshold range is measured, assisting the medical staff to respond quickly (Multimedia Appendix 3).

Mobile App

To enable efficient patient management and communication between the patients and medical staff, an Android-based mobile app was developed for patients in the LTSC. The app consisted of six functions: general guides for patients admitted to the LTSC, notice board, symptom questionnaire, vital sign reporting, question and answer (Q&A), and push reminders. The app was released on March 27, 2020 (Figure 5).

Figure 5. Interface of the mobile app for patients admitted to the LTCS. Q&A: question and answer.



General Guides for the LTSC

General information on admission to the LTSC and frequently asked questions by patients were organized for patient reference. Administrators could edit or register additional content directly through the app management web portal.

Notice Board

The notice board was used to send announcements to the patients. A push alarm was triggered when a new announcement was registered.

Symptom Questionnaire

Patients were asked to self-report their symptoms using the questionnaire in the mobile app. The questionnaire focused on the presence and frequency of symptoms most relevant to COVID-19 (cough, sputum, fever, rhinorrhea, sore throat, dyspnea, and chest pain) but also asked patients to report other more general symptoms (nausea, vomiting, abdominal discomfort, constipation, diarrhea, abdominal pain, insomnia, pain, and others). The self-reported data automatically populated the health record template.

Vital Sign Self-Reporting

In addition to the self-reporting of symptoms, patients were instructed to measure their vital signs twice a day, in the morning and afternoon, with the devices provided in each room and to enter the values into the mobile app. The vital sign measurements were also sent to the HIS to be reviewed by the medical staff.

Q&A

The Q&A function was a secure messaging platform for patients to submit questions to the medical team. The staff members replied through the app management web portal, and when a question was answered, a push notification was sent to the patient's mobile phone.

Push Alarms

Patients received a push alarm as a reminder to fill out the self-reporting questionnaires for symptoms and vital signs, when their question was answered, or when the medical staff posted a new announcement to the notice board.

Usability Survey

Twelve patients participated in the online surveys. The mean age of the respondents was 25 years (SD 6.25 years), and 7 (58%) were college students. For the mobile app, out of 5 points, the perceived usefulness showed the highest score at 4.62 points (SD 0.48), followed by satisfaction at 4.08 points (SD 1.41) and perceived ease of use at 3.81 points (SD 0.41) (Supplemental Table 1 in Multimedia Appendix 4). When patients rated the experience of using the wearable vital sign monitoring devices out of 5, perceived usefulness scored the highest at 4.45 points (SD 0.57), followed by perceived ease of use at 4.30 points (SD 0.59) and satisfaction at 3.98 points (SD 0.70) (Supplemental Table 2 in Multimedia Appendix 4).

A simple satisfaction survey regarding the usability of the mobile app was also conducted with the medical staff. Of the 24 participants who answered the questionnaire, 20 (83%) were nurses and 4 (17%) were physicians. The overall satisfaction score for the mobile app out of five points was 4.10 (SD 0.64) (Supplemental Table 1 in Multimedia Appendix 4).

Discussion

Principal Findings

The ICT-based patient management system introduced in this study was applied to the LTSC for clinically healthy COVID-19 patients operated by SNUH. Due to the infectious nature of the disease, it was difficult for medical staff to directly perform patient care; therefore, noncontact management was essential. SNUH introduced an ICT-based care system for the management of clinically healthy COVID-19 patients at the LTSC that fit the entire care flow, from preadmission and admission to discharge or transfer. The process started on March 1, 2020, and was completed in 4 weeks. Korea’s strategy of aggressive mitigation through early detection based on extensive diagnostic testing thorough epidemiological investigation and transparent management of COVID-19–related information has “flattened the curve” of the rate of infection [7]. The Ministry of Health and Welfare of Korea has tailored treatment policies and guidelines to match the disease profile of COVID-19 and the available medical resources. As part of their multifaceted approach to COVID-19, the Ministry relaxed regulations and temporarily approved the limited use of telemedicine and e-prescriptions [9]. Because direct telemedicine visits between health care professionals and patients were previously prohibited, the infrastructure and ecosystem around telemedicine were not fully developed [16]. Another approach was the decision to isolate patients with mild symptoms in designated LTSCs from March 1, 2020 [9]. The LTSCs were public or private facilities that were modified to accommodate and quarantine patients. SNUH, as a national hospital of Korea, was also asked to operate an LTSC at the SNUH Human Resources Development Center in Mungyeong. Most of the medical staff were located in the central monitoring center in Seoul and led patient care using telemedicine tools; meanwhile, some staff worked at the Mungyeong site in supportive roles and in preparation for emergencies.

The authors recognized the importance of a multidisciplinary approach to develop a system to support patient care and management at LTSCs as quickly as possible [17-19]. The Office of Hospital Information of SNUH, composed of several medical faculty members, nurses, pharmacists, medical record technicians, and IT experts, manages the IT services of the hospital and was well placed to lead this effort. Moreover, ezCaretech, the company that developed our HIS, is under contract to maintain the HIS and has a team located at SNUH; this team could make timely improvements to the system in a straightforward manner. In addition, HealthConnect, a health care application development company, has been working on a mobile app with SNUH for a project to collect patient-reported outcomes and enable viewing of the data in the HIS. This familiarity with the hospital interfaces and data structures enabled rapid development and deployment of the LTSC app.

Role of Health Care Information Exchange in the Pandemic Era

Health care information exchange between health care institutions plays an important role during pandemics and disasters [20-24]. In Korea, where the population has high accessibility to health care resources due to the national health insurance program, any patient can visit any hospital. Health records are fragmented among the health care institutions that the patient visits, as the EHR systems used by each institution are different and most are not interoperable [25]. To alleviate this problem, the Ministry of Health and Welfare initiated a nationwide healthcare information exchange (HIE) project in 2017. This service electronically and securely sends and receives medical records and image data of patients who have consented to provide personal information to medical institutions for patient care [26]. However, in the current HIE system, electronic information is generated only when the patient is referred to another medical institution; thus, it is difficult to use it in this situation. Furthermore, the project is not gaining traction as expected due to the cost burden to small and medium-sized hospitals to connect their EHRs with the national HIE system as well as the conflict between health care institutions and the government on reimbursement and penalty policies [27]. As most institutions were not prepared to share image data through the national HIE system, we developed a simple cloud-based image sharing platform to enable efficient exchange. This platform was immediately used for a 55-year-old woman who had no underlying disease and no symptoms at initial presentation but who progressed to pneumonia 3 days after admission. Her physician decided to transfer her to a local general hospital and shared her chest x-ray images through this system.

Lessons From the MERS Outbreak and Application for COVID-19 in South Korea

South Korea learned important lessons from its 2015 Middle East respiratory syndrome (MERS) outbreak, including the importance of increased transparency, early detection, and rapid diagnosis [28-30]. The Korea Centers for Disease Control and the Health Insurance Review & Assessment Service developed the international traveler information system and provides the service to all health care institutions in Korea [31]. When the patient has visited countries that are being screened, the international traveler information system sends out a real time alert to allow staff to recognize and prevent possible exposure. Similarly, to minimize exposure, the system was used to inform the physician or nurse whether the patient was arriving from the epicenter of the COVID-19 epidemic in Korea, Daegu and Gyeongsangbuk Province, since February 18, 2020.

Although the current pandemic is ongoing, we should also learn from this experience. Telemedicine is ideal for proper triage, early diagnosis, isolation, and treatment of patients [32]. Even in the United States, where access to telemedicine services, coverage, and reimbursement is available, telemedicine use is relatively uncommon [33]. Telemedicine use is even rarer in Korea, which is one of the few countries where patient-facing telemedicine is explicitly banned [16]. All stakeholders involved, including the government, medical communities, civic groups, and politicians, have different opinions. Some believe that Korea should approach this issue differently from other countries due to the nature of the medical system in Korea, which has excellent accessibility and low costs [34]. However, to limit community transmission during the COVID-19 pandemic, the Ministry of Health and Welfare has temporarily approved the limited use of telemedicine with e-prescriptions [9]. With recent studies reporting the advantages of proactive use of IT systems and strategies in infectious patients [35-37], we hope that this experience provides an impetus for the stakeholders of the health care system to improve and expand on the use and application of ICT in health care.

Conclusions

Faced with a novel infectious disease, we describe the implementation and experience of applying an ICT-based patient management system in the LTSC by SNUH. ICT-based tools and applications are increasingly gaining importance in health care, and we hope that our experience can provide insight into future technology-based infectious disease responses.

Acknowledgments

We would like to gratefully acknowledge our excellent team members, Suhyeon Kim, Jisu Ha, Buri Lee, and Sanghee Lim from the Hospital Information Development team, Myeonghee Song and Gyeongran Hong from the Medical Records team, Taewoo Kim and Donggyu Lee from the Hospital Information Strategy team, and Hyesook Lee from the Hospital Information System and Security team for their profound dedication and excellence. We would like to thank our colleagues Hyuck Kim from HealthConnect, Jinmo Kim and Sanghak Lee from Tribell Labs, and the IT specialists from ezCaretech. Above all, we would like to express deep gratitude to all experts and volunteers who are still fighting COVID-19. This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, South Korea (grant number: HI19C0790).

Authors' Contributions

KHK had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. YSB, KHK, CWJ, BC, MSK, and EK conceived and designed the study. YSB acquired, analyzed, and interpreted the data. YSB, SWC, and TK drafted the manuscript. Critical revision of the manuscript was provided by YSB and SWC. YSB and KHK provided administrative, technical, or material support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental Figure 1. Wearable vital sign recorder.
[PNG File , 280 KB - jmir_v22i6e19938_app1.png]

Multimedia Appendix 2
Supplemental Figure 2. Integration of data from the wearable vital sign device into the SNUH HIS.
[PNG File , 371 KB - jmir_v22i6e19938_app2.png]

Multimedia Appendix 3
Supplemental Figure 3. Central vital sign monitoring system.
[PNG File , 1485 KB - jmir_v22i6e19938_app3.png]

Multimedia Appendix 4
Supplemental Table 1. Survey on perceived usefulness, perceived ease of use, and satisfaction with the mobile app. Supplemental Table 2. Survey on perceived usefulness, perceived ease of use, and satisfaction with wearable devices by patients in the LTSC.

[DOCX File , 17 KB - jmir_v22i6e19938_app4.docx]

References

1. Paules CI, Marston HD, Fauci AS. Coronavirus Infections-More Than Just the Common Cold. JAMA 2020 Jan 23:Preprint. [doi: 10.1001/jama.2020.0757] [Medline: 31971553]
2. Kim D, Choe YJ, Jeong J. Understanding and Interpretation of Case Fatality Rate of Coronavirus Disease 2019. J Korean Med Sci 2020 Mar 30;35(12):e137 [FREE Full text] [doi: 10.3346/jkms.2020.35.e137] [Medline: 32233163]
3. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 — Studies Needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/nejmp2002125]
4. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
5. Korea Centers for Disease Control. Global Locations with COVID-19 URL: http://ncov.mohw.go.kr/en/bdBoardList.do?brdId=16&brdGubun=163&dataGubun=&ncvContSeq=&contSeq=&board_id=&gubun= [accessed 2020-03-28]
6. World Health Organization. Coronavirus disease (COVID-2019) situation reports URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-06-05]
7. Korea Centers for Disease Control. Coronavirus Disease-19, Republic of Korea URL: http://ncov.mohw.go.kr/en/ [accessed 2020-06-05]
8. Korean Society of Infectious Diseases, Korean Society of Pediatric Infectious Diseases, Korean Society of Epidemiology, Korean Society for Antimicrobial Therapy, Korean Society for Healthcare-associated Infection Control and Prevention, Korea Centers for Disease Control and Prevention. Report on the Epidemiological Features of Coronavirus Disease 2019 (COVID-19) Outbreak in the Republic of Korea from January 19 to March 2, 2020. J Korean Med Sci 2020 Mar 16;35(10):e112 [FREE Full text] [doi: 10.3346/jkms.2020.35.e112] [Medline: 32174069]
9. Ministry of Health and Welfare, Republic of Korea. 2020 Mar 09. Current Status of Response to COVID-19 and Future Plans URL: http://www.mofa.go.kr/viewer/skin/doc.html?fn=2020052006593622&rs=/viewer/result/202006 [accessed 2020-06-05]
10. Hick JL, Biddinger PD. Novel Coronavirus and Old Lessons — Preparing the Health System for the Pandemic. N Engl J Med 2020 May 14;382(20):e55. [doi: 10.1056/nejmp2005118]
11. Yoo S, Lee KH, Lee HJ, Ha K, Lim C, Chin HJ, et al. Seoul National University Bundang Hospital's Electronic System for Total Care. Healthc Inform Res 2012 Jun;18(2):145-152 [FREE Full text] [doi: 10.4258/hir.2012.18.2.145] [Medline: 22844650]
12. Ministry of Health and Welfare, Republic of Korea. 2020 Feb 04. Updates on COVID-19 in Republic of Korea (as of 20 April) URL: http://ncov.mohw.go.kr/en/tcmBoardView.do?brdId=12&brdGubun=125&dataGubun=&ncvContSeq=1974&contSeq=1974&board_id=&gubun= [accessed 2020-06-05]
13. Deshmukh V, Tripathi S, Patil A, Tripathy J. ScienceOpen Preprints. 2020 May 08. COVID-19 diagnostic multiplicity and its role in community surveillance and control URL: https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-.PPA94RC.v1 [accessed 2020-06-05]
14. Hong K, Lee S, Kim T, Huh H, Lee J, Kim S, et al. Guidelines for Laboratory Diagnosis of Coronavirus Disease 2019 (COVID-19) in Korea. Ann Lab Med 2020 Sep;40(5):351-360 [FREE Full text] [doi: 10.3343/alm.2020.40.5.351] [Medline: 32237288]
15. Korea Health Insurance Review and Assessment Service. 2020. COVID-19 Living and Treatment Support Center guidelines for medical reimbursement criteria and how to request it. Webpage in Korean URL: http://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020002000100&brdScnBltNo=4&brdBltNo=7876 [accessed 2020-04-03]
16. Oh J, Park Y, Jo EC, Kim S. Current Status and Progress of Telemedicine in Korea and Other Countries. Healthc Inform Res 2015 Oct;21(4):239-243 [FREE Full text] [doi: 10.4258/hir.2015.21.4.239] [Medline: 26618029]
17. Del Rio C, Malani PN. COVID-19-New Insights on a Rapidly Changing Epidemic. JAMA 2020 Feb 28:Preprint. [doi: 10.1001/jama.2020.3072] [Medline: 32108857]
18. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:Preprint. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
19. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Mar 03:Preprint. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
20. Lurie N, Dausey DJ, Knighton T, Moore M, Zakowski S, Deyton L. Community Planning for Pandemic Influenza: Lessons From the VA Health Care System. Disaster Med Public Health Prep 2013 Apr 08;2(4):251-257. [doi: 10.1097/dmp.0b013e31817dd143]
21. Boyd T, Savel T, Kesarinath G, Lee B, Stinn J. The Use of Public Health Grid Technology in the United States Centers for Disease Control and Prevention H1N1 Pandemic Response. 2010 Jan 01 Presented at: 24th IEEE International Conference on Advanced Information Networking and Applications Workshops; April 10-23, 2010; Perth, Western Australia. [doi: 10.1109/waina.2010.60]

22. Martin R, Conseil A. Public health policy and law for pandemic influenza: a case for European harmonization? J Health Polit Policy Law 2012 Dec;37(6):1091-1110. [doi: 10.1215/03616878-1813854] [Medline: 22899835]
23. Hammond WE, Bailey C, Boucher P, Spohr M, Whitaker P. Connecting information to improve health. Health Aff (Millwood) 2010 Feb;29(2):284-288. [doi: 10.1377/hlthaff.2009.0903] [Medline: 20348075]
24. Han SH, Lee MH, Kim SG, Jeong JY, Lee BN, Choi MS, et al. Implementation of Medical Information Exchange System Based on EHR Standard. Healthc Inform Res 2010 Dec;16(4):281-289 [FREE Full text] [doi: 10.4258/hir.2010.16.4.281] [Medline: 21818447]
25. Moon TJ. Light and Shadows of the Korean Healthcare System. J Korean Med Sci 2012;27(Suppl):S3. [doi: 10.3346/jkms.2012.27.s.s3]
26. Korea Social Security Information Service. 2019 Apr 01. Healthcare Information Exchange. Webpage in Korean URL: http://www.ssis.or.kr/lay1/S1T760C797/contents.do [accessed 2020-06-08]
27. Choi S. Korea Doctor's News. 2019 Aug 07. Korea medical association requests that the hospital information exchange project be discontinued without assistance. Article in Korean URL: https://www.doctorsnews.co.kr/news/articleView.html?idxno=130423 [accessed 2020-06-08]
28. Petersen E, Hui DS, Perlman S, Zumla A. Middle East Respiratory Syndrome - advancing the public health and research agenda on MERS - lessons from the South Korea outbreak. Int J Infect Dis 2015 Jul;36:54-55 [FREE Full text] [doi: 10.1016/j.ijid.2015.06.004] [Medline: 26072036]
29. Khan A, Farooqui A, Guan Y, Kelvin DJ. Lessons to learn from MERS-CoV outbreak in South Korea. J Infect Dev Ctries 2015 Jul 04;9(6):543-546 [FREE Full text] [doi: 10.3855/jidc.7278] [Medline: 26142661]
30. Lim P. Middle East respiratory syndrome (MERS) in Asia: lessons gleaned from the South Korean outbreak. Trans R Soc Trop Med Hyg 2015 Sep;109(9):541-542 [FREE Full text] [doi: 10.1093/trstmh/trv064] [Medline: 26286944]
31. Korea Centers for Disease Control. 2018 Sep 28. International Traveler Information System (ITS) URL: http://www.cdc.go.kr/gallery.es?mid=a20503010000&bid=0002&act=view&list_no=140757 [accessed 2020-06-08]
32. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
33. Barnett ML, Ray KN, Souza J, Mehrotra A. Trends in Telemedicine Use in a Large Commercially Insured Population, 2005-2017. JAMA 2018 Nov 27;320(20):2147-2149 [FREE Full text] [doi: 10.1001/jama.2018.12354] [Medline: 30480716]
34. Chang H. Evaluation Framework for Telemedicine Using the Logical Framework Approach and a Fishbone Diagram. Healthc Inform Res 2015 Oct;21(4):230-238 [FREE Full text] [doi: 10.4258/hir.2015.21.4.230] [Medline: 26618028]
35. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20:1357633X2091656. [doi: 10.1177/1357633x20916567]
36. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for COVID-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
37. Arsenault M, Evans B, Karanofsky M, Gardie J, Shulha M. Canadian Family Physician. 2020 Mar 26. Covid 19 - Practising Telemedicine in the pandemic URL: https://www.cfp.ca/news/2020/03/26/3-26-1 [accessed 2020-06-05]

Abbreviations

CMS: central monitoring system
COVID-19: coronavirus disease
DICOM: Digital Imaging and Communications in Medicine
EHR: electronic health record
e-prescription: electronic prescription
HIE: healthcare information exchange
HIS: hospital information system
ICT: information and communications technology
IT: information technology
LTSC: living and treatment support center
MERS: Middle East respiratory syndrome
PACS: picture archiving and communication system
Q&A: question and answer
RT-PCR: real time polymerase–chain reaction
SNUH: Seoul National University Hospital

Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by HA Park, R Lee, T Jamieson; comments to author 26.05.20; revised version received 02.06.20; accepted 02.06.20; published 12.06.20.

Please cite as:
Bae YS, Kim KH, Choi SW, Ko T, Jeong CW, Cho B, Kim MS, Kang E
Information Technology–Based Management of Clinically Healthy COVID-19 Patients: Lessons From a Living and Treatment Support Center Operated by Seoul National University Hospital
J Med Internet Res 2020;22(6):e19938
URL: http://www.jmir.org/2020/6/e19938/
doi:10.2196/19938
PMID:32490843

Viewpoint

Why African Americans Are a Potential Target for COVID-19 Infection in the United States

Alireza Hamidian Jahromi[1], MD, MRCS; Anahid Hamidianjahromi[2], MD

[1]Department of Plastic Surgery, The University of Tennessee Health Science Center, Memphis, TN, United States
[2]Jahrom University of Medical Sciences, Jahrom, Iran

Corresponding Author:
Alireza Hamidian Jahromi, MD, MRCS
Department of Plastic Surgery
The University of Tennessee Health Science Center
910 Madison Avenue, Suite 315
Memphis, TN, 38103
United States
Phone: 1 318 518 4600
Email: alirezahamidian@yahoo.com

Abstract

Since the World Health Organization declared the coronavirus disease (COVID-19) outbreak a pandemic, significant changes have occurred in the United States as the infection spread reached and passed its exponential phase. A stringent analysis of COVID-19 epidemiologic data requires time and would generally be expected to happen with significant delay after the exponential phase of the disease is over and when the focus of the health care system is diverted away from crisis management. Although much has been said about high-risk groups and the vulnerability of the elderly and patients with underlying comorbidities, the impact of race on the susceptibility of ethnic minorities living in indigent communities has not been discussed in detail worldwide and specifically in the United States. There are currently some data on disparities between African American and Caucasian populations for COVID-19 infection and mortality. While health care authorities are reorganizing resources and infrastructure to provide care for symptomatic COVID-19 patients, they should not shy away from protecting the general public as a whole and specifically the most vulnerable members of society, such as the elderly, ethnic minorities, and people with underlying comorbidities.

(J Med Internet Res 2020;22(6):e19934) doi:10.2196/19934

KEYWORDS

coronavirus; COVID-19; African American; mortality; race; virus; minority; infectious disease

Although it has only been a short time since March 11th, 2020, when the World Health Organization officially announced that the coronavirus disease (COVID-19) outbreak had reached the level of a pandemic [1], things have changed significantly for the United States. The country has reached and passed the exponential phase of the COVID-19 pandemic, with over 1.7 million confirmed cases and close to 105,000 COVID-19–related deaths [2]. These numbers seem to reflect only the tip of the iceberg; the figures would be much higher if we are to include asymptomatic carriers and mildly symptomatic untested cases in the population. As shown in a study by Ing et al [3], testing all of the passengers of a cruise ship revealed a ratio of asymptomatic COVID-19–positive patients to symptomatic patients to be 4:1. Furthermore, a second wave of COVID-19 infection is always a possibility.

A stringent analysis of COVID-19 epidemiologic data requires time and would generally be expected to happen with some delay after the exponential phase of the disease is over and when the focus of the health care system is not on crisis management. Although much is said about high-risk groups and the vulnerability of the elderly and patients with underlying comorbidities (eg, hypertension, diabetes, malignancies, cerebrovascular disease, and underlying pulmonary/cardiovascular disease like chronic obstructive pulmonary disease and asthma [4-6]), the impact of race on the susceptibility of ethnic minorities living in indigent communities has not been discussed in detail worldwide and in the United States. Myers [7] and Dyer [8] brought up this issue and discussed as to why, in their opinion, ethnic minorities, in particular African American or black populations, are more vulnerable to COVID-19 infection and mortality. Dyer [8] also brought up the same issue with other “invisible communities” (ie, Navajo Nations communities of Arizona, New Mexico, and Utah and “undocumented immigrants”) and proposed delayed health care seeking behavior, inferior care, difficulty in trusting authorities (eg, fear of charges against those without legal immigration documents upon receiving care or being tested for COVID-19) as possible explanations. Dyer [8] believes that African Americans are generally suspicious of the health care system following the Tuskegee syphilis scandal (1932-1972), where black participants with a syphilis diagnosis were monitored for almost 40 years and left untreated despite the availability of treatment for syphilis at the time; this mistrust continues to impact medical advice–seeking behaviors among black populations in the United States [8].

Knowledge of comorbid risk factors as well as vulnerable groups can be useful for clinicians in directing early and appropriate medical management and allocating resources when handling patients with COVID-19. As “ethnicity” is still considered a sensitive/controversial topic in the United States, its inclusion as a research topic as well as patients’ demographic data should be cautiously evaluated; even the Centers for Disease Control and Prevention does not include race and ethnicity information in their publicly available data when presenting the demographic and geographic distributions of COVID-19 cases. Myers [7] provided invaluable, practical recommendations to health care authorities and hospital administrations caring for vulnerable populations with COVID-19, including patients with comorbidities and ethnic minorities. Cooney [9] presented examples from Illinois, Michigan, and Milwaukee County in Wisconsin to show disparities and the overrepresentation of COVID-19 cases and mortalities among African Americans compared to other ethnic groups, especially Caucasians. Although these areas do not constitute the majority of the United States, these are a handful of exceptional areas where the local government agencies have included race as part of the shared information when presenting the most vulnerable and affected segments of society “hardest hit by the virus” [9,10]. While African Americans in Illinois, as an example, account for 17% of confirmed COVID-19 cases and 29.7% of attributed deaths, they make up only 15% of the state’s population [10]. In Louisiana, based on the 2019 US census, 33% of the population is African American, but this group accounts for 55% of COVID-19–related deaths, according to data from the Department of Health [11]. North Carolina, Michigan, St Louis, New York City, and Alabama are other examples of disparities in COVID-19–related deaths and ethnicity, with higher death rates reported in African Americans compared to Caucasians [12-15].

This disparity is not a confirmed fact across the board, but it is present in some states as mentioned above. There are some possible explanations for this pattern. The proposed association between COVID-19 and angiotensin-converting enzyme II (ACE-II) secondary to ACE-II being a co-receptor for respiratory viral entry, along with potential innate ACE-II receptor differences in function, response to medications, and gene polymorphism in the African American population compared to other ethnic groups [16,17] could be an explanation that needs further investigation. Having higher incidence of hypertension, diabetes, malignancies, heart failure, cerebrovascular disease, and underlying pulmonary/cardiovascular disease could also make the African American community a potential high-risk group for COVID-19 [7,8,18,19]. Higher BMI among African Americans compared to other groups, particularly Caucasians, has been shown to be present in children and adults [20,21]. This contributes to a general decrease in self-rated health among older African Americans and can even interfere with the protective effects of education and income [22]. Higher average BMI in African Americans could potentially decrease their survival rate in cases with COVID-19–related acute respiratory distress syndrome and increase the rate of associated complications [23]. Other possible contributing factors to COVID-19 susceptibility among African Americans include lower socioeconomic status, crowded living conditions, living in densely populated parts of cities, reduced access to hygienic products and personal protection equipment, unemployment, employment in the more virus-exposed fields of the job market, along with cultural habits that hinder medical advice seeking (although their exact role warrants further investigations) [24]. Myers [7] went a step further and proposed the influence of “systemic racism on baseline health” as a compounding factor influencing the health of ethnic minorities including African American people. While health care authorities are reorganizing resources and infrastructure to provide care for symptomatic COVID-19 patients [25], they should not forgo protecting the general public as a whole and specifically the most vulnerable members of society, such as the elderly, ethnic minorities, and people with underlying comorbidities. We agree with Myers [7] and Fouad et al [26] that the following are essential steps toward a fairer and more effective health care system when battling COVID-19 infection: collecting data on patient care during the COVID-19 era including race, implementation of social policies to reduce economic burden, culturally sensitive and appropriate messaging tailored to certain communities, increased accessibility of testing and vaccination (after vaccine development) for vulnerable communities, real-time reporting of disparities and reflective redistribution of resources, aggressive treatment of comorbidities simultaneously while treating COVID-19 infections, and prioritization of community-based health care systems that primarily serve ethnic minorities. We can turn this crisis into an opportunity to reduce disparities in care in our health care system [26].

Acknowledgments

The authors would like to thank Ms Talicia Tarver, who provided us with editorial assistance.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2020 Mar 11. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-05-29]
2. Worldometer. 2020 Jun 08. COVID-19 Coronavirus Pandemic URL: https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?, [accessed 2020-05-28]
3. Ing AJ, Cocks C, Green JP. COVID-19: in the footsteps of Ernest Shackleton. Thorax 2020 May 27:thoraxjnl-2020-215091. [doi: 10.1136/thoraxjnl-2020-215091] [Medline: 32461231]
4. Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. Aging (Albany NY) 2020 Apr 08;12(7):6049-6057 [FREE Full text] [doi: 10.18632/aging.103000] [Medline: 32267833]
5. Shahid Z, Kalayanamitra R, McClafferty B, Kepko D, Ramgobin D, Patel R, et al. COVID-19 and Older Adults: What We Know. J Am Geriatr Soc 2020 May 20;68(5):926-929 [FREE Full text] [doi: 10.1111/jgs.16472] [Medline: 32255507]
6. The Lancet Oncology. COVID-19: global consequences for oncology. The Lancet Oncology 2020 Apr;21(4):467. [doi: 10.1016/s1470-2045(20)30175-3]
7. Myers E. Compounding Health Risks and Increased Vulnerability to SARS-CoV-2 for Racial and Ethnic Minorities and Low Socioeconomic Status Individuals in the United States. Preprints.org 2020 Apr 15:e [FREE Full text] [doi: 10.20944/preprints202004.0234.v1]
8. Dyer O. Covid-19: Black people and other minorities are hardest hit in US. BMJ 2020 Apr 14;369:m1483. [doi: 10.1136/bmj.m1483] [Medline: 32291262]
9. Cooney E. Stat. 2020 Apr 06. ‘We’re flying blind’: African Americans may be bearing the brunt of Covid-19, but access to data is limited URL: https://www.statnews.com/2020/04/06/flying-blind-african-americans-disparities-covid-19-data-limited/ [accessed 2020-05-29]
10. Illinois Department of Public Health. 2020. COVID-19 Statistics URL: https://www.dph.illinois.gov/covid19/covid19-statistics [accessed 2020-05-29]
11. Louisiana Department of Health. 2020. COVID-19 URL: http://ldh.la.gov/coronavirus/ [accessed 2020-05-29]
12. NYC Health. 2020. COVID-19: Data URL: https://www1.nyc.gov/site/doh/covid/covid-19-data.page [accessed 2020-05-29]
13. Echols F. The St. Louis American. 2020 Apr 18. All 12 COVID-19 deaths in the City of St. Louis were black URL: http://www.stlamerican.com/your_health_matters/covid_19/all-12-covid-19-deaths-in-the-city-of-st-louis-were-black/article_da7ed56c-79d1-11ea-85bc-7b8539eaf346.html [accessed 2020-05-29]
14. Alabama Public Health. 2020. Characteristics of Laboratory-Confirmed Cases of COVID-19 URL: https://www.alabamapublichealth.gov/infectiousdiseases/assets/cov-al-cases-041220.pdf [accessed 2020-05-29]
15. NCDHHS COVID-19 Response. 2020. Cases URL: https://covid19.ncdhhs.gov/dashboard/cases [accessed 2020-05-29]
16. Patel AB, Verma A. COVID-19 and Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers: What Is the Evidence? JAMA 2020 Mar 24:e. [doi: 10.1001/jama.2020.4812] [Medline: 32208485]
17. Cohall D, Ojeh N, Ferrario CM, Adams OP, Nunez-Smith M. Is hypertension in African-descent populations contributed to by an imbalance in the activities of the ACE2/Ang-(1-7)/Mas and the ACE/Ang II/AT axes? J Renin Angiotensin Aldosterone Syst 2020 Feb 22;21(1):1470320320908186 [FREE Full text] [doi: 10.1177/1470320320908186] [Medline: 32089050]
18. Zilbermint M, Hannah-Shmouni F, Stratakis C. Genetics of Hypertension in African Americans and Others of African Descent. Int J Mol Sci 2019 Mar 02;20(5):1081 [FREE Full text] [doi: 10.3390/ijms20051081] [Medline: 30832344]
19. Clements JM, West BT, Yaker Z, Lauinger B, McCullers D, Haubert J, et al. Disparities in diabetes-related multiple chronic conditions and mortality: The influence of race. Diabetes Res Clin Pract 2020 Jan;159:107984. [doi: 10.1016/j.diabres.2019.107984] [Medline: 31846667]
20. Anderson PM, Butcher KF, Schanzenbach DW. Understanding recent trends in childhood obesity in the United States. Econ Hum Biol 2019 Aug;34:16-25. [doi: 10.1016/j.ehb.2019.02.002] [Medline: 30910341]
21. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of Obesity Among Adults and Youth: United States, 2015-2016. NCHS Data Brief 2017 Oct(288):1-8 [FREE Full text] [Medline: 29155689]
22. Assari S, Cobb S, Bazargan M. Race by Gender Differences in the Protective Effects of Education and Income Against Subsequent Changes in Self-rated Health, Physical Activity, and Body Mass Index Among Older Americans. J Health Econ Dev 2019;1(2):9-21 [FREE Full text] [Medline: 32201864]
23. Ball L, Serpa Neto A, Pelosi P. Obesity and survival in critically ill patients with acute respiratory distress syndrome: a paradox within the paradox. Crit Care 2017 May 03;21(1):114 [FREE Full text] [doi: 10.1186/s13054-017-1682-5] [Medline: 28532465]
24. American Psychological Association. 2020. Ethnic and Racial Minorities & Socioeconomic Status URL: https://www.apa.org/pi/ses/resources/publications/minorities [accessed 2020-05-29]
25. Hamidian JA, Mazloom S, Ballard D. What the European and American health care systems can learn from China COVID-19 epidemic; action planning using purpose designed medical telecommunication, courier services, homebased quarantine, and COVID-19 walk-in centers. Immunopathol Persa 2020 2020 Mar;6(2):e17 [FREE Full text]

26. Fouad MN, Ruffin J, Vickers SM. COVID-19 is Out of Proportion in African Americans. This Will Come as No Surprise…. Am J Med 2020 May 19:e [FREE Full text] [doi: 10.1016/j.amjmed.2020.04.008] [Medline: 32442510]

Abbreviations

ACE-II: angiotensin-converting enzyme II
COVID-19: coronavirus disease

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by D Ballard, D Laborde, N Yamada; comments to author 28.05.20; revised version received 29.05.20; accepted 03.06.20; published 12.06.20.

Please cite as:
Hamidian Jahromi A, Hamidianjahromi A
Why African Americans Are a Potential Target for COVID-19 Infection in the United States
J Med Internet Res 2020;22(6):e19934
URL: http://www.jmir.org/2020/6/e19934/
doi:10.2196/19934
PMID:32496205

Original Paper

Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo

Zhuo-Ying Tao[1,2], BDS; Guang Chu[2], BDS; Colman McGrath[3], BDS, MSc, PhD; Fang Hua[4], BDS, MDS, PhD; Yiu Yan Leung[1], BDS, MDS, PhD; Wei-Fa Yang[1], BDS, MDS; Yu-Xiong Su[1], BDS, MD, PhD

[1]Division of Oral and Maxillofacial Surgery, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, Research Center of Stomatology, Xi'an Jiaotong University College of Stomatology, Xi'an, Shaanxi Province, China

[3]Applied Oral Sciences & Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China (Hong Kong)

[4]Center for Evidence-Based Stomatology, School & Hospital of Stomatology, Wuhan University, Wuhan, Hubei Province, China

Corresponding Author:
Yu-Xiong Su, BDS, MD, PhD
Division of Oral and Maxillofacial Surgery
Faculty of Dentistry
The University of Hong Kong
34 Hospital Road
Hong Kong,
China (Hong Kong)
Phone: 852 28590267
Email: richsu@hku.hk

Related Articles:

Comment in: https://www.jmir.org/2021/5/e26255

Comment in: https://www.jmir.org/2021/5/e29145

Abstract

Background: Social media has become increasingly important as a source of information for the public and is widely used for health-related information. The outbreak of the coronavirus disease (COVID-19) has exerted a negative impact on dental practices.

Objective: The aim of this study is to analyze the nature and diffusion of COVID-19–related oral health information on the Chinese social media site Weibo.

Methods: A total of 15,900 tweets related to oral health and dentistry information from Weibo during the COVID-19 outbreak in China (December 31, 2019, to March 16, 2020) were included in our study. Two researchers coded 1000 of the total tweets in advance, and two main thematic categories with eight subtypes were refined. The included tweets were analyzed over time and geographic region, and coded into eight thematic categories. Additionally, the time distributions of tweets containing information about dental services, needs of dental treatment, and home oral care during the COVID-19 epidemic were further analyzed.

Results: People reacted rapidly to the emerging severe acute respiratory syndrome coronavirus 2 threat to dental services, and a large amount of COVID-19–related oral health information was tweeted on Weibo. The time and geographic distribution of tweets shared similarities with epidemiological data of the COVID-19 outbreak in China. Tweets containing home oral care and dental services content were the most frequently exchanged information (n=4803/15,900, 30.20% and n=4478, 28.16%, respectively). Significant differences of public attention were found between various types of bloggers in dental services–related tweets ($P<.001$), and the tweets from the government and media engaged the most public attention. The distributions of tweets containing information about dental services, needs of dental treatment, and home oral care information dynamically changed with time.

Conclusions: Our study overviewed and analyzed social media data on the dental services and oral health information during the COVID-19 epidemic, thus, providing insights for government organizations, media, and dental professionals to better facilitate oral health communication and efficiently shape public concern through social media when routine dental services are unavailable during an unprecedented event. The study of the nature and distribution of social media can serve as a useful adjunct tool to help make public health policies.

(J Med Internet Res 2020;22(6):e19981) doi:10.2196/19981

KEYWORDS

COVID-19; dentistry; oral health; online health; social media; tweet; Weibo; China; health information

Introduction

The outbreak of the coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first identified from Wuhan, Hubei Province, China, has almost swept across the whole world and constituted a public health emergency of “pandemic” proportions [1]. SARS-CoV-2 is the third zoonotic human coronavirus emerging in this century after severe acute respiratory syndrome–related coronavirus and Middle East respiratory syndrome–related coronavirus (MERS-CoV) [2]. As a global pandemic, COVID-19 has affected people from more than 200 countries and regions, leading to 2,719,897 laboratory-confirmed cases and 187,705 deaths as of April 25, 2020 [3]. The total cases outside of China has outnumbered China more than 20 times. Strict prevention measures and effective therapeutics are urgently needed for the control of the pandemic.

COVID-19 has posed a particular threat to the practice of dentistry. With the identification of SARS-CoV-2 in the saliva of patients who are infected [4] and the transmission of COVID-19 through asymptomatic carriers [5-7] or potential patients in incubation status [8], awareness of the risk of COVID-19 spreading during dental procedures was of considerable concern among dental professionals [9-11]. Moreover, droplets and aerosols generated from high-speed dental handpieces, ultrasonic instruments, or 3-way syringes has the potential for direct or indirect cross transmissions of coronavirus between patients and dental care providers [12]. Therefore, during the early and outbreak stages of the COVID-19 epidemic in China, a majority of dental practices were suspended and most of the routine dental services were not available for the public, resulting in the inconvenience for potential patients to seek dental treatment. During this special period, social media played an important role in the exchange of dentistry- and oral health–related information among the public.

The internet, especially social media, is becoming increasingly important as a source of information for public health issues since it provides free and immediate access to large volumes of data [13]. In the past decade, social media has not only changed the pattern of spread for health-related information and the communication mode between patients and health care providers, but also drawn great attraction from researchers to study the distribution of diseases [14], the diffusion of health-related information and misinformation [15,16], the public reactions to health events [17-19], and more. Accumulating studies using social media for health care research have been published annually, providing insights for public health surveillance or helping develop health policy [14,20-23]. During the COVID-19 epidemic, the social media search index was identified as a promising predictor of new cases of COVID-19 infections [24].

Sina Weibo, similar to Twitter, is the most popular online microblog platform in China. Weibo allows its users to tweet or retweet messages optionally with links, pictures, or videos attached. The public reactions of Chinese people to the MERS-CoV and H7N9 outbreaks were significantly strong on Weibo [17]. Considering that the COVID-19 outbreak has exerted an impact on dental practices, we proposed the following research question: how was dentistry- or oral health–related information during the COVID-19 epidemic “tweeted” and communicated about on Weibo? The aim of this study was to investigate the nature and diffusion of COVID-19–related dentistry or oral health information on Weibo and determine the public reactions to tweets with this content, thus, providing an overview and reflection of the supply and demand of dental services under the epidemic on social media.

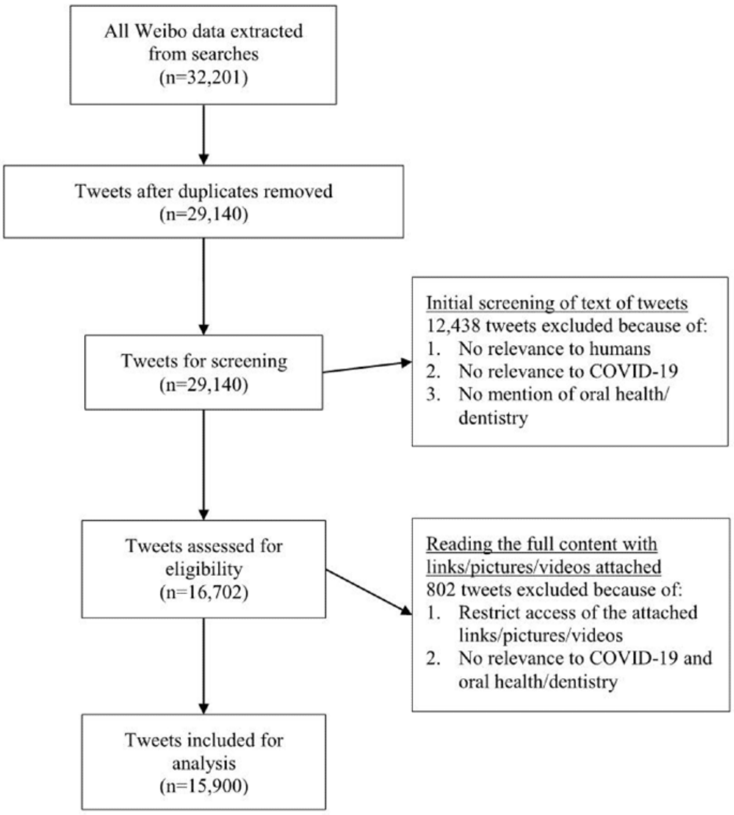
Methods

Study Design and Search Strategy

A study of “COVID-19-related oral health information” tweets on Sina Weibo was performed. A new anonymous Weibo account was created with only the name, gender, and date of birth provided upon registration. Using a new account without search histories, previous likes, or friends can avoid preferential links promoted by Weibo. Four keywords related to COVID-19 (pneumonia of unknown cause, coronavirus, COVID-19, and epidemic) and two keywords for dentistry (stomatology and dentistry) in Chinese characters were employed to search tweets about COVID-19 and dentistry or oral health on Weibo. Eight independent searches with a combination of one keyword for COVID-19 and the other for dentistry were carried out on March 17, 2020, through the new account. We selected December 31, 2019, as the start date of tweets since, on this day, the pneumonia of unknown cause (the name of COVID-19 at the time) in Wuhan was officially reported to the World Health Organization, and the first group of epidemiologists were dispatched by the Chinese Center for Disease Control and Prevention (CCDC) to support the control of this emerging infectious disease (EID) in Wuhan.

Figure 1 is the flow diagram used in our study. In total, 32,201 tweets in the first 77 days (December 31, 2019, to March 16, 2020) of the COVID-19 epidemic were identified. All Weibo data, including the full-text content; post time; numbers of likes, shares, and comments; and blogger information (name, ID, homepage website, number of followers, and personal introduction) of each tweet were extracted with a Python-based platform *Gooseeker*, which could retrieve all the tweets from each independent research through the account we created. A total of 29,140 tweets after removal of duplicates were scanned to exclude non–human-related and irrelevant results. After screening, there were 16,702 tweets included for further eligibility assessment, namely, reading the full content and the links, pictures, or videos attached to tweets as well. Restricted access of links, pictures, or videos were also excluded since only the information available for the public was assessed. Finally, the remaining 15,900 tweets were included for further analysis. The initial screening work and eligibility assessment was conducted by two researchers (ZT and GC) together.

Figure 1. Flow chart of the study. COVID-19: coronavirus disease.



Coding Procedure

Two researchers (ZT and GC) with expertise in dentistry completed the coding. First, the two coders were asked to pilot the project by coding 1000 (6.28%) of the total 15,900 tweets to develop and refine the coding schemes for thematic categories. Two main thematic categories for all tweets were determined initially: COVID-19–related and oral health–related information. The information related to COVID-19 was further subtyped into five domains: epidemiology, pathology, symptoms, diagnosis, and prevention; the information related to oral health was subtyped into three domains: dental services, needs of dental treatment, and home oral care information. The tweets that were related to oral health and COVID-19 but inappropriate to be sorted into any of the categories were labelled “others.” The definition and examples of each category are shown in Multimedia Appendix 1.

Second, to test the feasibility and reliability of categories, a weighed Kappa test was used to assess interrater and intrarater agreement of coding between two researchers. Two researchers were asked to classify 200 randomly selected tweets and reclassify 2 weeks after the first coding. The results of the weighed Kappa test of the two researchers were 0.983 for interrater agreement and 0.994 (ZT) and 0.983 (GC) for intrarater agreement, which indicated excellent reliability of coding procedure.

Third, after agreement for the coding of the tweets was confirmed, all 15,900 tweets were randomly separated into two groups and classified by two researchers. The tweets were coded to more than one category if containing miscellaneous information. Tweets in each thematic category were divided into two types according to the numbers of followers (less than 1000 followers and 1000 or more followers).

Postanalysis of the Included Tweets

The tweets of each day were counted and compared with COVID-19 daily new cases and deaths in China for time distribution analysis. Among 15,900 tweets, 1682 tweets with location information were analyzed for geographic distribution and compared with the regional distribution of total COVID-19 cases by March 16, 2020. The epidemiological data were obtained from the official website of the CCDC [25]. In addition, the tweets and retweets conveying similar information were extracted and grouped to analyze the topics of interest. The mean and SD of the numbers of likes, shares, and comments in each group were calculated.

Meanwhile, the tweets related to oral health information were further analyzed. Specifically, the time distribution of tweets containing information about the risks of COVID-19 transmission during dental procedures, notices of stopping all or part of dental services, need for dental treatment, home oral care, protective measures during dental services, and notices of restoring dental services during the COVID-19 epidemic were analyzed. The public needs for dental treatment and home oral care information tweeted on Weibo were further categorized and counted. The Kruskal-Wallis test was used to determine the differences of public reactions (numbers of likes, shares, and comments) to tweets from different types of bloggers (governments, media, dental clinics or hospitals, dentists or dental nurses, online health platforms, and others). In addition, the Mann-Whitney U test was used to compare the public reactions to tweets related to oral health information tweeted by the same types of bloggers.

All statistical analyses were carried out with Microsoft Excel (Microsoft Corporation) and SPSS software 18.0 (SPSS Inc), and *P*<.05 was considered significant.

Results

Time and Geographic Distribution of Weibo Tweets

As shown in Figure 2, the time distribution of the Weibo data was similar to the distribution of new cases and deaths during this period of time. From December 31, 2019, to March 16, 2020, the daily count of tweets in Weibo was low during the first 20 days, then increased with fluctuations starting on January 20, 2020, presented three major peaks before mid-February, and finally gradually decreased and kept at a steady low level until March 16, 2020. The first peak on January 22, 2020, was between two milestone events of the COVID-19 outbreak in China, namely, official confirmation of human-to-human transmission of COVID-19 on January 20 and the start of the lockdown in Wuhan on January 23, 2020. The second peak on January 28, 2020, was composed of tweets and retweets refuting the misinformation for COVID-19 prevention. The highest peak of 1316 tweets on February 8, 2020, followed the death of Dr Wenliang Li, one of the first doctors who flagged the new coronavirus outbreak and raised alarms to the public. This peak occurred because retweets of the news that aerosol acts as a transmission route for COVID-19 as disseminated in a press conference of the Shanghai government office at 2 PM that day.

Further analyses were conducted on 1682 tweets whose geographic distribution could be identified and compared to the regional distribution of total cases. The geographic location was optional for the users when posting tweets on Weibo. Therefore, only 1682 tweets included geographic location information. As demonstrated in Figure 3, the provinces (Zhejiang, Jiangsu, Guangdong, Shandong, Sichuan, and Henan) that witnessed the most tweets (>100) had more than 500 COVID-19 cases, higher than most of the other provinces in China. In northwest China, with no more than 100 confirmed cases in each province, tweets were less frequently distributed. More tweets were posted from central and eastern coastal provinces. However, tweets posted from Hubei, which was the original epicenter of the COVID-19 outbreak and had the most cumulative number of cases, surprisingly, were no more than 100.

Figure 2. Time distribution of tweets and new cases and deaths of COVID-19 in China. COVID-19: coronavirus disease.

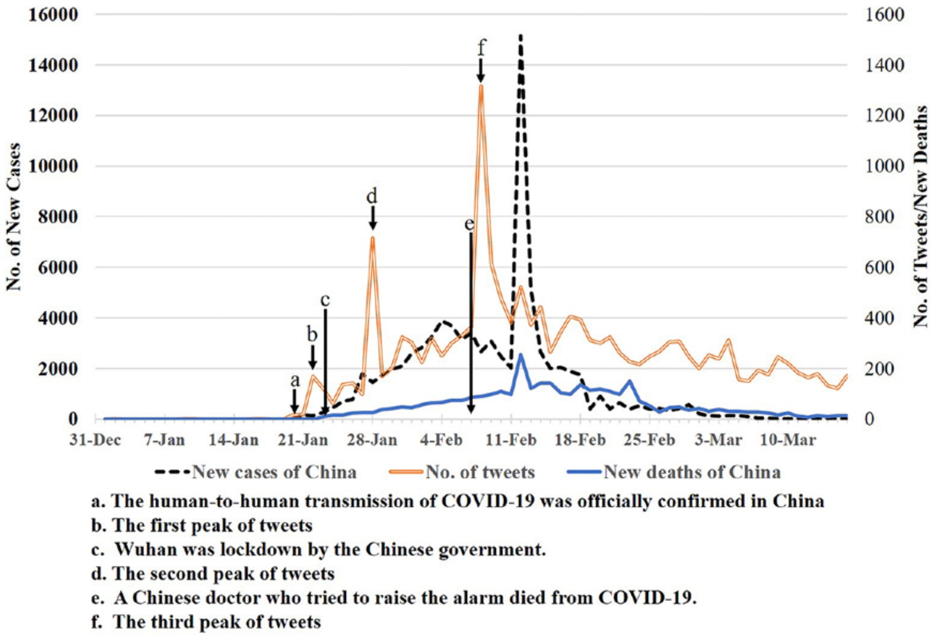
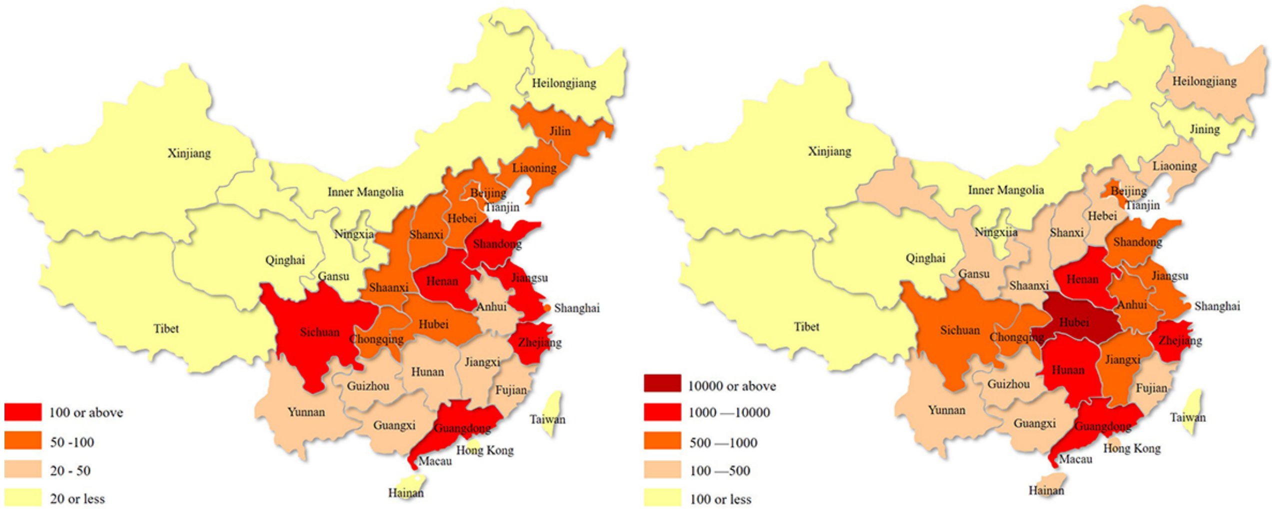


Figure 3. The geographic distributions of Weibo tweets (left) and total coronavirus disease cases (right) in the region.



Thematic Distribution of Tweets and Highly Retweeted Information

Among 15,900 tweets included in our study, 79.81% (n=12,690) were oral health–related information and 38.86% (n=6180) contained background knowledge of COVID-19. As shown in Figure 4, the most commonly exchanged types of tweets related to oral health was home oral care information (n=4803, 30.20%), followed by dental service–related information (n=4478, 28.16%), the need for dental treatment (n=2793, 17.57%) during the epidemic, and other information about oral health or dentistry and COVID-19 (n=616, 3.87%). In terms of background knowledge about COVID-19, information about the prevention of COVID-19 (n=3404, 21.40%) witnessed the most tweets and retweets on Weibo, and epidemiology of COVID-19 (n=2390, 15.04%) was also common, while only a relatively small proportion of tweets mentioned aspects of pathology and symptoms (n=139, 0.87%) or diagnosis (n=247, 1.54%). The background information of COVID-19 (except the pathology and symptoms) was mostly tweeted by bloggers with more than 1000 followers. The dental services and home oral care information were highly tweeted by bloggers with 1000 or more followers, whereas the need for dental treatment were tweeted mostly by bloggers with followers less than 1000 (Figure 4).

Some information was frequently tweeted or retweeted on Weibo and the top five pieces of widely diffused information were selected for evaluating the public reactions (Table 1). The most tweeted information was the news that aerosol acts as a transmission route of COVID-19, officially announced by the government office of Shanghai on February 8, 2020. There were 1406 tweets or retweets of this news or related aerosol information, and it gained 321.28 likes, 24.85 shares, and 19.70 comments on average. The second most tweeted information related to the risks of COVID-19 spread in dental clinics due to the aerosol generation, widely exchanged by dentists and dental clinics and hospitals, which averagely garnered 4.72 likes, 1.92 shares, and 1.31 comments. The next 3 groups of tweets were refutations of three pieces of widely spread misinformation of prevention measures for the coronavirus. As shown in Figure 5, the first misinformation that gargling with saltwater or mouthwash can prevent COVID-19 by lowering the level of coronavirus in the saliva spread among the public at an early stage of the epidemic (around the end of January). Soon the refutation of this misinformation was tweeted by the official platform “Weibo Refutes Rumours” and retweeted by others much more than the misinformation itself. A similar pattern was seen in another 2 groups of misinformation, namely, that eating garlic or applying oral spray or disinfectants can kill the coronavirus in oral cavities to protect from COVID-19 infection (Figure 5).

Figure 4. Thematic distributions of tweets with COVID-19–related oral health information. COVID-19: coronavirus disease.

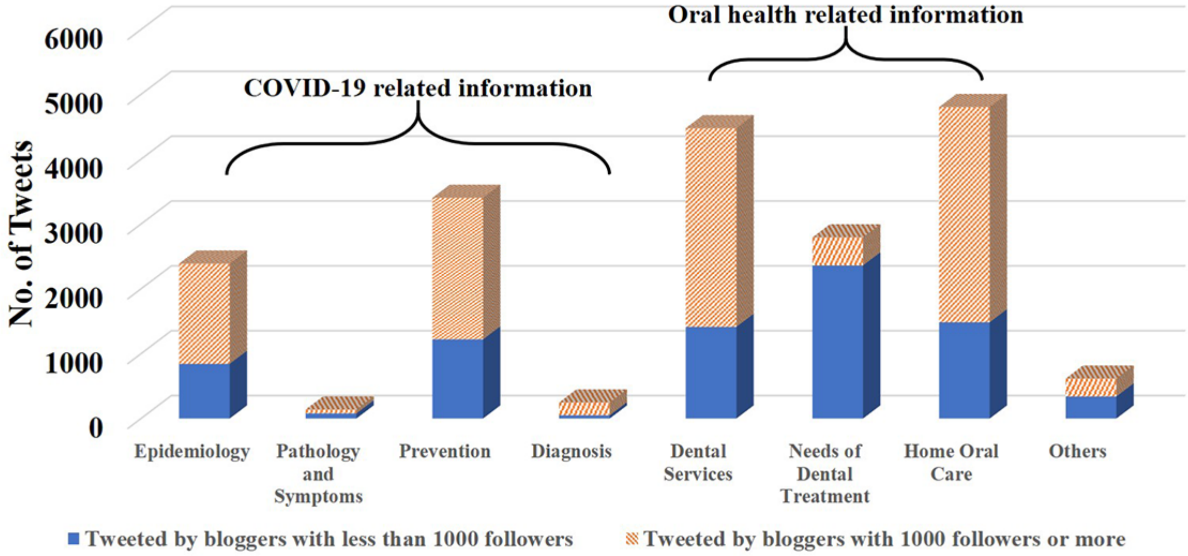
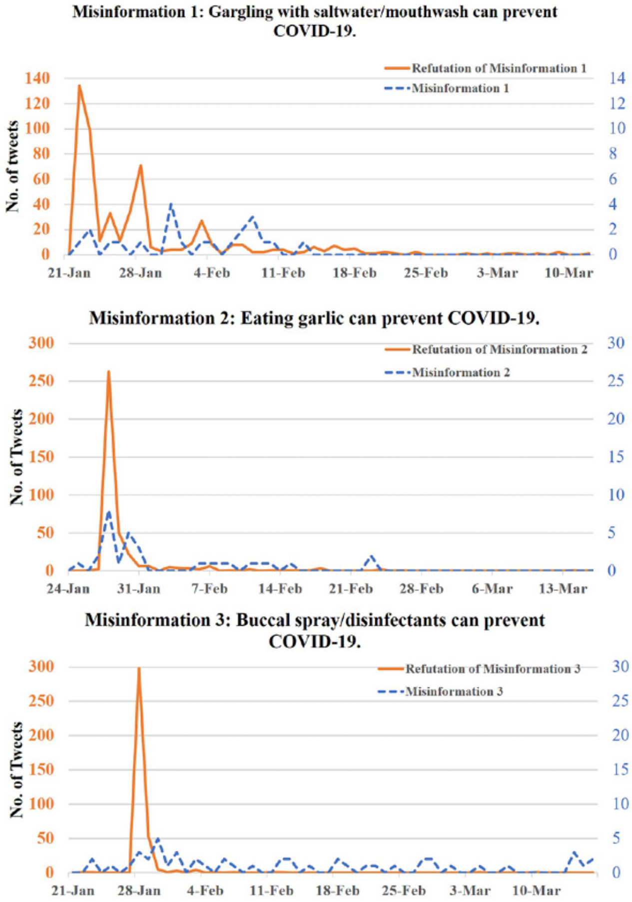


Table 1. Public reactions to highly tweeted information on Weibo.

Most highly tweeted information	Count, n	Public reactions		
		Likes, mean (SD)	Shares, mean (SD)	Comments, mean (SD)
The news propagandizing aerosol as a transmission route of COVID-19[a]	1406	321.28 (11,225.52)	24.85 (581.45)	19.70 (539.29)
Risks of COVID-19 spread by dental clinics due to the aerosol created by dental handpieces	659	4.72 (56.91)	1.92 (14.34)	1.31 (7.29)
Refutation of the misinformation that gargling with saltwater or mouthwash can prevent COVID-19	468	105.99 (1010.50)	27.27 (312.07)	9.98 (62.31)
Refutation of the misinformation that eating garlic can kill the novel coronavirus in the oral cavity	389	50.25 (639.83)	17.62 (237.45)	7.60 (84.71)
Refutation of the misinformation that oral spray/disinfectants can prevent COVID-19.	372	0.25 (1.88)	0.43 (2.74)	0.05 (0.32)

[a] COVID-19: coronavirus disease.

Figure 5. The time distributions of tweets related to misinformation for COVID-19 prevention and its refutations. COVID-19: coronavirus disease.



Misinformation 1: Gargling with saltwater/mouthwash can prevent COVID-19.

Misinformation 2: Eating garlic can prevent COVID-19.

Misinformation 3: Buccal spray/disinfectants can prevent COVID-19.

Disruption of Dental Services During the Epidemic

Tweets with different types of oral health–related information were distributed differently during the COVID-19 epidemic. As shown in Figure 6, the risk of COVID-19 spread during dental procedures was proposed during the early stage of the epidemic, and the associated tweets and retweets peaked on January 25 and 26, 2020, after the start of the Wuhan lockdown on January 23. During the first half of February, under the first level emergency response to COVID-19 in China, the government, public hospitals, and private dental clinics delivered notices to stop all or part of dental services to the public. Followed by this was the peak of information about home oral health care in the second half of February. From late February to early March, tweets associated with notices of restoring dental services and protective measures during dental treatment gradually increased due to the control of COVID-19 in China.

When the dental services were not available from the end of January to early March, many bloggers complained of oral problems and sought dental care on Weibo. There were a steady number of tweets (around 40-80 tweets/day; Figure 6) during this time period. As shown in Table 2 those complaining of toothaches or wisdom tooth problems (eg, pericoronitis, decay) were most common, followed by orthodontic-related problems, oral ulcers, and pediatric oral diseases; oral cancer and implants or prostheses only occupied a small proportion. Additionally, the rest of the tweets were related to seeking oral care for other dental diseases including tooth decay and gingival bleeding or without specific reasons.

Interestingly, numerous tweets with home oral care content were found on Weibo when the majority of dental services were not available for the public, much more than the tweets seeking oral health care as previously mentioned (4803 vs 2973 tweets). Information about daily oral care, how to deal with dental emergencies at home, and online consultation services shared similar proportions among these tweets (n=1684/4803, 35.06%; n=2092, 43.56%; and n=2029, 42.24%, respectively; Figure 7). Notably, 21.50% (n=362/1684) of daily oral care–related tweets presented commercial advertisements of oral hygiene products including toothbrushes, toothpastes, and dental floss.

As for the public responses to dental services and home oral care–related tweets from different types of bloggers, the number of likes, shares, and comments for tweets from governments, media, dental clinics and hospitals, dentists and dental nurses, online health platforms, and other nondental bloggers were counted and analyzed (only bloggers with >1000 followers were included; Table 3). Significant differences of public attention were found between various types of bloggers in dental service–related tweets (χ^2_4=113.883, 99.037, 49.544 for numbers of likes, shares, and comments, respectively, and P<.001 for all of them). The same findings were seen in tweets with home oral care content (χ^2_5=292.817, 186.265, 264.250, respectively, and P<.001 for numbers of likes, shares, and comments). The governments and media tweeted a large number of tweets related to dental services but fewer related to home oral care content (492 vs 98 tweets from governments, 805 vs 87 tweets from media, respectively) and received public responses with high average numbers of likes, shares, and comments in both categories. The government-generated tweets with dental service information received significantly more shares and comments but less likes than with home oral care content (dental services vs home oral care, *P*<.001 for numbers of likes, shares, and comments), while the media-tweeted information related to dental services garnered much less public attention compared to home oral care content (*P*<.001 for numbers of likes, shares, and comments). As for the tweets from dental clinics and hospitals, the public seemed to pay more attention to the tweets associated with home oral care information than dental services (*P*=.001 for numbers of likes and comments), though tweets in both categories were intensive. Dentists and dental nurses tweeted a mass of information related to oral health, and there were no significant differences between public responses for tweets with home oral care information and dental services (*P*=.37, *P*=.30, and *P*=.81 for numbers of likes, shares, and comments, respectively). Notably, nearly one-fifth of tweets with home oral care information were provided by online health platforms and garnered 4.05 likes, 1.28 shares, and 0.59 comments on average.

Figure 6. Time distributions of oral health–related tweets during the COVID-19 epidemic. COVID-19: coronavirus disease.

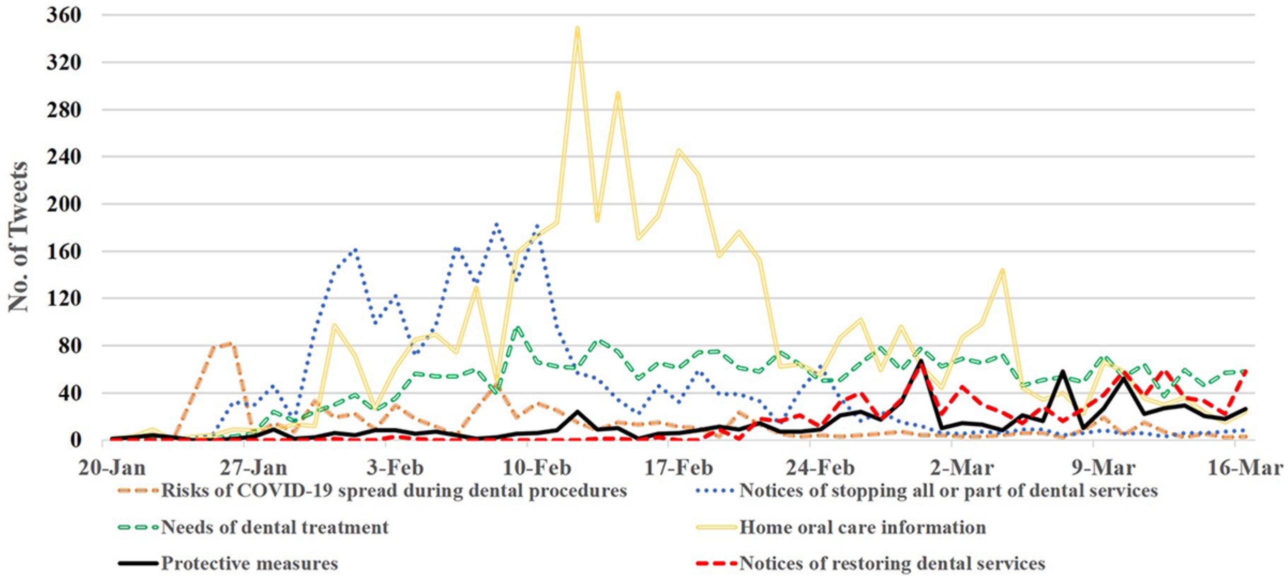


Table 2. Needs for dental treatment during COVID-19 epidemic.

Needs of dental treatment	Number of tweets (n=2793), n (%)[a]
Toothache or wisdom tooth problem	1132 (40.53)
Oral ulcer	264 (9.45)
Orthodontic problem	536 (19.19)
Implants or prostheses	31 (1.11)
Pediatric oral diseases	81 (2.90)
Oral cancer	41 (1.47)
Others or not specific	788 (28.21)

[a] The sum value of all parts is over 100% because some tweets mentioned more than one need of dental treatment.

Figure 7. Thematic distributions of tweets with home oral care information.

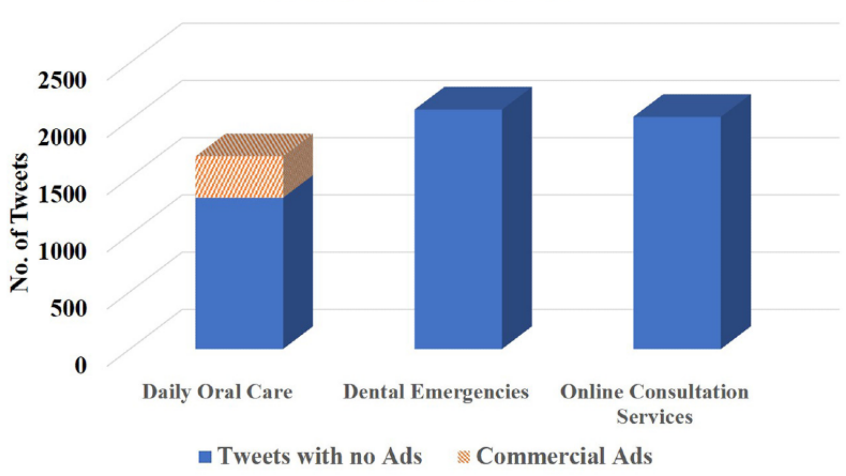


Table 3. Comparison of public reactions to tweets with dental services and home oral care information from different types of bloggers.

Public reactions	Blogger categories					
	Governments	Media	Dental clinics/hospitals	Dentists/dental nurses	Online health platform	Others
Count, n (%)						
DS[a]	492 (16.05)	805 (26.26)	563 (18.37)	363 (11.84)	0 (0)	842 (24.47)
HOC[b]	98 (2.95)	87 (2.62)	1049 (31.60)	874 (26.33)	660 (19.88)	551 (16.6)
Likes						
DS, mean (SD)	11.91 (68.00)	54.36 (1053.84)	1.66 (8.21)	2.47 (47.20)	N/A[c]	11.47 (78.58)
HOC, mean (SD)	14.35 (128.38)	168.10 (1110.52)	1.27 (11.09)	26.82 (377.06)	4.05 (64.39)	5.69 (38.82)
Z value	–6.858	–5.49	–3.259	–0.896	N/A	–14.75
P value	<.001	<.001	.001	.37	N/A	<.001
Shares						
DS, mean (SD)	2.75 (8.44)	4.63 (33.37)	0.88 (5.75)	0.49 (9.37)	N/A	2.37 (14.81)
HOC, mean (SD)	1.94 (8.27)	32.48 (182.30)	0.88 (6.75)	5.34 (42.89)	1.28 (17.34)	3.81 (49.31)
Z value	–5.943	–5.234	–0.565	–1.045	N/A	–10.06
P value	<.001	<.001	.57	.30	N/A	<.001
Comments						
DS, mean (SD)	3.97 (15.79)	7.43 (74.11)	0.70 (2.59)	0.42 (8.08)	N/A	4.15 (23.85)
HOC, mean (SD)	2.87 (23.22)	25.36 (154.88)	0.33 (1.56)	5.55 (51.73)	0.59 (8.85)	2.39 (12.42)
Z value	–5.925	–3.81	–3.279	–0.236	N/A	–12.68
P value	<.001	<.001	.001	.81	N/A	<.001

[a] DS: dental services.

[b] HOC: home oral care.

[c] Not applicable.

Discussion

Principal Results

Since the COVID-19 outbreak exerted a negative impact on dental practices [12], the Chinese online community reacted rapidly and tweeted considerable information associated with COVID-19 and dentistry and oral health. This is the first study to analyze COVID-19–related oral health information that was presented on Chinese social media and evaluate the public interaction with this information. As shown in our study, COVID-19–related oral health content tweeted on Weibo may serve as a key indicator of the supply and demand of dental services under the epidemic in China.

The Distribution of Tweets and the Epidemiology of COVID-19

It was interesting to note that the time distribution of the Weibo data was approximately consistent with the trend of both daily new cases and deaths. Our finding is similar to a previous study on H7N9–related tweets on Weibo, which identified a positive correlation between the number of daily tweets and the cumulative case fatality rate of H7N9 [26]. Accordingly, the time distribution of social media may help predict the trend of new cases and deaths during the epidemic, thus, serving as a cost-effective and useful tool for epidemiology study.

Additionally, the peaks of tweets were influenced by some milestone events of COVID-19, including the official confirmation of human-to-human transmission of COVID-19, lockdown of Wuhan, and the death of Dr Wenliang Li. The possible explanation for this phenomenon is that social media engages more public attention during outbreaks of EIDs, especially when important news about the epidemic is released [27]. Similarly, during the H1N1 pandemic in 2009 and the outbreaks of MERS-CoV in 2012 and H7N9 in 2013, the time distribution of related tweets on social media were impacted by milestone events, such as the official announcement of the first case diagnosed and the level of the epidemic [17,28].

Although our results show that the geographical distribution of the tweets was roughly consistent with the distribution of total COVID-19 cases, Hubei Province was an exception. As the first epicenter of a newly identified infectious disease in the world, people may have been overwhelmed, and the concerns for oral health problems and demands for dental services were unavoidably suppressed. This phenomenon should attract more attention from public health policy makers.

Social media should not only provide true and useful information for the public, but also possess self-correction function for misinformation [29]. As illustrated in our study, during the early stage of the COVID-19 outbreak, some misleading information for prevention of COVID-19 was diffused among the online community, which attributed to a natural fear for the unknown and unexpectable disaster. This phenomenon is quite common during outbreaks of emerging diseases. When yellow fever re-emerged during 2015-2017, nearly two-thirds of tweets associated with yellow fever contained misinformation including some improper treatments [30]. After the first Zika infection case confirmed in the United States, grassroots users on Twitter amplified social concerns and even tweeted conspiracy theories [31]. A systemic review about social media and outbreaks of EIDs showed that 20%-30% of the EID–related YouTube videos contained inaccurate or misleading information [27]. In our study, the refutations of misinformation were tweeted by the official platform “Weibo Refutes Rumours” and immediately exchanged by other users with a larger scale on Weibo. This may contribute to calming the public panic with scientific knowledge.

Dental Services and Dental Care Needs During the COVID-19 Outbreak

Social media can amplify the spread of contents compared to traditional mass media [32]. In this study, the most widely diffused information was about aerosol as a transmission route of COVID-19 and its effect on dental practices, which was first tweeted on January 25, 2020, on Weibo (earlier than scientific publications). The study that first confirmed SARS-CoV-2 existed in saliva was published on February 12, 2020 [4], and the study that first announced the possibility of COVID-19 transmission by aerosol-generating dental procedures was published on February 20, 2020 [9]. The updates and spread of information on social media are updated to the minute, and the communication on social media is in real time, much faster than traditional media and online scientific publications. Moreover, increasing numbers of bloggers in public health professions are active on social media and tweet scientific knowledge on epidemics of infectious diseases, which plays a crucial role in shaping public awareness and response to an emerging disaster [33].

The analysis of the COVID-19–related oral health contents on Weibo provided an overview of the supply and demand of dental services during the COVID-19 outbreak. The public concern about the risks of spreading coronavirus by aerosols was disseminated earlier than the official notices of stopping part of or all dental services. With insufficient dental services, the public needs for oral care could not be satisfied, and an increased number of users complained of oral diseases or sought for consultation on Weibo. Our study showed dynamic changes of information related to supply and demand of dental services during this period, indicating that social media can serve as a useful tool for the monitoring of medical and health demands during an unprecedented time [34].

Social media plays an increasingly important role in health policy making [20,21,35]. As shown in our review, the urge of stopping dental services to avoid potential risks of coronavirus transmission between patients and dental care providers was tweeted by dental professionals and retweeted by other bloggers at the end of January, which may contribute to the public health policy about suspension of dental practices from the governmental agencies at the beginning of February. This important function of social media benefits from the multidirectional conduit of social networks, leading to more efficient and wide diffusion of information on social media than on traditional media [36] and more frequent interactions between individuals and public health organizations and policy makers [37]. Regarding social media as a rapidly maturing channel of communication, the policy makers can obtain evidence to help make health policy decisions, disseminate the policies on social media, and monitor the public reactions to the policies [38].

In response to perceived unmet dental care needs from the online community, home oral care information was highly tweeted on Weibo to satisfy the public needs for both daily oral care and dental emergencies. Remote dental consultations were also achieved through Weibo, not only providing diagnosis and suggestions for patients with oral health problems but also avoiding risks of coronavirus transmission. This correlation between dental service disruption and an increased use of social media as a means of communication is not uncommon during a disaster or emergency [39]. Previous studies have also recommended social media for dental public health surveillance due to its potential in monitoring episodes of dental pain [40,41]. The active exchange of information related to oral care and the interaction between patients and dental professionals were found from our Weibo data, presenting a promising communication mode between health care providers and patients as a supplement for the traditional doctor-patient relationship.

The Variation of Public Reactions to Tweets From Different Bloggers

In our study, the online community reacted variously to the tweets from different bloggers. The tweets posted by governments and media attracted more responses from the public due to their authority [42]. The government and media should take the responsibility to tweet true and real time information. Surprisingly, the influence of dental clinics and hospitals, and dentists and dental nurses was not strong enough despite tweeting high-quality information of oral health. The important role of the emerging online health platforms was identified in this study. The online health platforms with more followers than individual dental professionals can transmit health-related information more efficiently and benefit more people [43]. From the perspective of social equality, the services from online health platforms can also cover those hard-to-reach populations, making it an equitable access to health care for the public. Therefore, our results suggest that the online health platform has potential to be a promising solution for cost-efficient health care information and medical consultations.

Limitations

Several limitations of our study should be acknowledged. First, there was inevitable bias of information in the data collection process. For example, Weibo is more popular among young people rather than the aged and is more accessible for economically developed regions. Therefore, the impact of age and regional distribution of users may need to be considered when interpreting the results. As a real time social media, some tweets were deleted by bloggers and some bloggers' accounts were suspended by Sina Weibo, which led to direct information loss. In the geographic analysis of tweets, around 90% did not provide location information, thus affecting the overall objectivity and accuracy of results to a certain extent. Second, our study did not provide any data on the characteristics of Weibo users who viewed and shared these tweets, and the complete diffusion route of tweets were not extracted and analyzed. Therefore, the audience of the information and the diffusion scale of tweets could not be accurately evaluated. Third, the information provided by online consultation services was not available, and thus, the quality of online consultations and the effects on the patients remained unknown.

Conclusions

To the best of our knowledge, this is the first study to comprehensively overview and analyze social media data on the dental services and oral health information during the COVID-19 epidemic in China. Based on our results, it is evident that social media users reacted immediately to the emerging SARS-CoV-2 threat to dental practices. Social media not only contributed to public health surveillance and policy making but also served as a bridge between oral health information providers and the patients. The findings illustrate the relationship between social media information with the supply and demand of dental services during the outbreak of the COVID-19 epidemic in China. In addition, the study provides insights for government organizations, media, and dental professionals to efficiently affect and shape public awareness, and disseminate dental public health information through social media.

Acknowledgments

This work was supported by the University of Hong Kong research output prize 2019.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Thematic distribution of tweets related to oral health/dentistry during the coronavirus disease epidemic, December 31, 2019, to March 16, 2020.

[DOCX File , 21 KB - jmir_v22i6e19981_app1.docx]

References

1. Mahase E. Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. BMJ 2020 Mar 12;368:m1036. [doi: 10.1136/bmj.m1036] [Medline: 32165426]

2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020 Feb 22;395(10224):565-574 [FREE Full text] [doi: 10.1016/S0140-6736(20)30251-8] [Medline: 32007145]
3. World Health Organization. 2020. Coronavirus disease 2019 (COVID-19) situation report – 96 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200425-sitrep-96-covid-19.pdf?sfvrsn=a33836bb_4 [accessed 2020-04-25]
4. To K, Tsang O, Chik-Yan Yip C, Chan K, Wu T, Chan J, et al. Consistent detection of 2019 novel coronavirus in saliva. Clin Infect Dis 2020 Feb 12:1-3 [FREE Full text] [doi: 10.1093/cid/ciaa149] [Medline: 32047895]
5. Liu Y, Liao C, Chang C, Chou C, Lin Y. A locally transmitted case of SARS-CoV-2 infection in Taiwan. N Engl J Med 2020 Mar 12;382(11):1070-1072 [FREE Full text] [doi: 10.1056/NEJMc2001573] [Medline: 32050059]
6. Bai Y, Yao L, Wei T, Tian F, Jin D, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA 2020 Feb 21:1406 [FREE Full text] [doi: 10.1001/jama.2020.2565] [Medline: 32083643]
7. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med 2020 Mar 05;382(10):970-971 [FREE Full text] [doi: 10.1056/NEJMc2001468] [Medline: 32003551]
8. Yu P, Zhu J, Zhang Z, Han Y. A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. J Infect Dis 2020 May 11;221(11):1757-1761 [FREE Full text] [doi: 10.1093/infdis/jiaa077] [Medline: 32067043]
9. Sabino-Silva R, Jardim ACG, Siqueira WL. Coronavirus COVID-19 impacts to dentistry and potential salivary diagnosis. Clin Oral Investig 2020 Apr;24(4):1619-1621 [FREE Full text] [doi: 10.1007/s00784-020-03248-x] [Medline: 32078048]
10. Meng L, Hua F, Bian Z. Coronavirus disease 2019 (COVID-19): emerging and future challenges for dental and oral medicine. J Dent Res 2020 May;99(5):481-487 [FREE Full text] [doi: 10.1177/0022034520914246] [Medline: 32162995]
11. Ather A, Patel B, Ruparel NB, Diogenes A, Hargreaves KM. Coronavirus disease 19 (COVID-19): implications for clinical dental care. J Endod 2020 May;46(5):584-595 [FREE Full text] [doi: 10.1016/j.joen.2020.03.008] [Medline: 32273156]
12. Peng X, Xu X, Li Y, Cheng L, Zhou X, Ren B. Transmission routes of 2019-nCoV and controls in dental practice. Int J Oral Sci 2020 Mar 03;12(1):9 [FREE Full text] [doi: 10.1038/s41368-020-0075-9] [Medline: 32127517]
13. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
14. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
15. Chen L, Wang X, Peng T. Nature and diffusion of gynecologic cancer-related misinformation on social media: analysis of tweets. J Med Internet Res 2018 Oct 16;20(10):e11515 [FREE Full text] [doi: 10.2196/11515] [Medline: 30327289]
16. Ho A, McGrath C, Mattheos N. Social media patient testimonials in implant dentistry: information or misinformation? Clin Oral Implants Res 2017 Jul;28(7):791-800. [doi: 10.1111/clr.12883] [Medline: 27279455]
17. Fung IC, Fu K, Ying Y, Schaible B, Hao Y, Chan C, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. Infect Dis Poverty 2013 Dec 20;2(1):31 [FREE Full text] [doi: 10.1186/2049-9957-2-31] [Medline: 24359669]
18. Fung IC, Cai J, Hao Y, Ying Y, Chan BSB, Tse ZTH, et al. Global Handwashing Day 2012: a qualitative content analysis of Chinese social media reaction to a health promotion event. Western Pac Surveill Response J 2015;6(3):34-42 [FREE Full text] [doi: 10.5365/WPSAR.2015.6.2.003] [Medline: 26668765]
19. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
20. Chen X, Lun Y, Yan J, Hao T, Weng H. Discovering thematic change and evolution of utilizing social media for healthcare research. BMC Med Inform Decis Mak 2019 Apr 09;19(Suppl 2):50 [FREE Full text] [doi: 10.1186/s12911-019-0757-4] [Medline: 30961624]
21. Huber J, Woods T, Fushi A, Duong M, Eidelman A, Zalal A, et al. Social media research strategy to understand clinician and public perception of health care messages. JDR Clin Trans Res 2020 Jan;5(1):71-81 [FREE Full text] [doi: 10.1177/2380084419849439] [Medline: 31067411]
22. Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, et al. Mining the characteristics of COVID-19 patients in China: analysis of social media posts. J Med Internet Res 2020 May 17;22(5):e19087 [FREE Full text] [doi: 10.2196/19087] [Medline: 32401210]
23. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]
24. Qin L, Sun Q, Wang Y, Wu K, Chen M, Shia B, et al. Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. Int J Environ Res Public Health 2020 Mar 31;17(7):2365 [FREE Full text] [doi: 10.3390/ijerph17072365] [Medline: 32244425]

25. Chinese Center for Disease Control and Prevention. 2020. The distribution of COVID-19 pandemic URL: http://2019ncov.chinacdc.cn/2019-nCoV/ [accessed 2020-03-16]
26. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
27. Tang L, Bie B, Park S, Zhi D. Social media and outbreaks of emerging infectious diseases: a systematic review of literature. Am J Infect Control 2018 Sep;46(9):962-972 [FREE Full text] [doi: 10.1016/j.ajic.2018.02.010] [Medline: 29628293]
28. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
29. Vijaykumar S, Nowak G, Himelboim I, Jin Y. Managing social media rumors and misinformation during outbreaks. Am J Infect Control 2018 Jul;46(7):850. [doi: 10.1016/j.ajic.2018.03.014] [Medline: 29661633]
30. Ortiz-Martínez Y, Jiménez-Arcia LF. Yellow fever outbreaks and Twitter: rumors and misinformation. Am J Infect Control 2017 Jul 01;45(7):816-817. [doi: 10.1016/j.ajic.2017.02.027] [Medline: 28343702]
31. Vijaykumar S, Nowak G, Himelboim I, Jin Y. Virtual Zika transmission after the first U.S. case: who said what and how it spread on Twitter. Am J Infect Control 2018 May;46(5):549-557. [doi: 10.1016/j.ajic.2017.10.015] [Medline: 29306490]
32. Margolis MA, Brewer NT, Shah PD, Calo WA, Gilkey MB. Stories about HPV vaccine in social media, traditional media, and conversations. Prev Med 2019 Jan;118:251-256. [doi: 10.1016/j.ypmed.2018.11.005] [Medline: 30414396]
33. Lowrey W, Evans W, Gower KK, Robinson JA, Ginter PM, McCormick LC, et al. Effective media communication of disasters: pressing problems and recommendations. BMC Public Health 2007 Jun 06;7:97 [FREE Full text] [doi: 10.1186/1471-2458-7-97] [Medline: 17553153]
34. Lackie E, Fairchild A. The birth control pill, thromboembolic disease, science and the media: a historical review of the relationship. Contraception 2016 Oct;94(4):295-302. [doi: 10.1016/j.contraception.2016.06.009] [Medline: 27343747]
35. Yeung D. Social media as a catalyst for policy action and social change for health and well-being: viewpoint. J Med Internet Res 2018 Mar 19;20(3):e94 [FREE Full text] [doi: 10.2196/jmir.8508] [Medline: 29555624]
36. Obar JA, Wildman S. Social media definition and the governance challenge: an introduction to the special issue. Telecommunications Policy 2015 Oct;39(9):745-750 [FREE Full text] [doi: 10.1016/j.telpol.2015.07.014]
37. Gatewood J, Monks SL, Singletary CR, Vidrascu E, Moore JB. Social media in public health: strategies to distill, package, and disseminate public health research. J Public Health Manag Pract 2019 Oct 22:1-4. [doi: 10.1097/PHH.0000000000001096] [Medline: 31651729]
38. Roland D. Social media, health policy, and knowledge translation. J Am Coll Radiol 2018 Jan;15(1 Pt B):149-152. [doi: 10.1016/j.jacr.2017.09.009] [Medline: 29122510]
39. Bosch D. Use of social media and e-Government in disasters: 2016 Louisiana floods case study. J Emerg Manag 2017;15(6):391-405. [doi: 10.5055/jem.2017.0347] [Medline: 29308600]
40. Eke PI. Using social media for research and public health surveillance. J Dent Res 2011 Sep;90(9):1045-1046 [FREE Full text] [doi: 10.1177/0022034511415277] [Medline: 21768305]
41. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
42. Yun GW, Morin D, Park S, Joa CY, Labbe B, Lim J, et al. Social media and flu: media Twitter accounts as agenda setters. Int J Med Inform 2016 Jul;91:67-73. [doi: 10.1016/j.ijmedinf.2016.04.009] [Medline: 27185510]
43. Terrasse M, Gorin M, Sisti D. Social media, e-Health, and medical ethics. Hastings Cent Rep 2019 Jan;49(1):24-33. [doi: 10.1002/hast.975] [Medline: 30790306]

Abbreviations

CCDC: Chinese Center for Disease Control and Prevention
COVID-19: coronavirus disease
EID: emerging infectious disease
MERS-CoV: Middle East respiratory syndrome–related coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 08.05.20; peer-reviewed by N Andiesta, MA Bahrami, C Basch, R Zowalla; comments to author 25.05.20; revised version received 03.06.20; accepted 04.06.20; published 15.06.20.

Please cite as:
Tao ZY, Chu G, McGrath C, Hua F, Leung YY, Yang WF, Su YX
Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo
J Med Internet Res 2020;22(6):e19981
URL: http://www.jmir.org/2020/6/e19981/
doi:10.2196/19981
PMID:32501808

Original Paper

Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States

Siyue Li[1], PhD; Bo Feng[2], PhD; Wang Liao[2], PhD; Wenjing Pan[3], PhD

[1]College of Media and International Culture, Zhejiang University, Hangzhou, China
[2]Department of Communication, University of California at Davis, Davis, CA, United States
[3]School of Journalism and Communication, Renmin University of China, Beijing, China

Corresponding Author:
Wenjing Pan, PhD
School of Journalism and Communication
Renmin University of China
507, School of Journalism and Communication
59 Zhongguancun St, Haidian District
Beijing, 100872
China
Phone: 86 010 82500855
Email: wenjingpan@ruc.edu.cn

Abstract

Background: During the coronavirus disease (COVID-19) pandemic, engagement in preventive behaviors and getting tested for the virus play a crucial role in protecting people from contracting the new coronavirus.

Objective: This study aims to examine how internet use, risk awareness, and demographic characteristics are associated with engagement in preventative behaviors and testing during the COVID-19 pandemic in the United States.

Methods: A cross-sectional survey was conducted on Amazon Mechanical Turk from April 10, 2020, to April 14, 2020. Participants' internet use (in terms of the extent of receiving information pertaining to COVID-19), risk awareness (whether any immediate family members, close friends or relatives, or people in local communities tested positive for COVID-19), demographics (sex, age, ethnicity, income, education level, marital status, and employment status), as well as their engagement in preventative behaviors and testing were assessed.

Results: Our data included 979 valid responses from the United States. Participants who received more COVID-19–related health information online reported more frequent effort to engage in all types of preventive behaviors: wearing a facemask in public (odds ratio [OR] 1.55, 95% CI 1.34-1.79, *P*<.001), washing hands (OR 1.58, 95% CI 1.35-1.85, *P*<.001), covering nose and mouth when sneezing and coughing (OR 1.78, 95% CI 1.52-2.10, *P*<.001), keeping social distance with others (OR 1.41, 95% CI 1.21-1.65, *P*<.001), staying home (OR 1.40, 95% CI 1.20-1.62, *P*<.001), avoiding using public transportation (OR 1.57, 95% CI 1.32-1.88, *P*<.001), and cleaning frequently used surfaces (OR 1.55, 95% CI 1.34-1.79, *P*<.001). Compared with participants who did not have positive cases in their social circles, those who had immediate family members (OR 1.48, 95% CI 8.28-26.44, *P*<.001) or close friends and relatives (OR 2.52, 95% CI 1.58-4.03, *P*<.001) who tested positive were more likely to get tested. Participants' sex, age, ethnicity, marital status, and employment status were also associated with preventive behaviors and testing.

Conclusions: Our findings revealed that the extent of receiving COVID-19–related information online, risk awareness, and demographic characteristics including sex, ethnicity, age, marital status, and employment status are key factors associated with US residents' engagement in various preventive behaviors and testing for COVID-19.

(J Med Internet Res 2020;22(6):e19782) doi:10.2196/19782

KEYWORDS

COVID-19; coronavirus; preventive behaviors; testing; online health information; risk awareness

Introduction

Since its initial outbreak in late December 2019, the coronavirus disease (COVID-19) pandemic has caused and is continuing to cause a severe, large-scale impact on individuals and societies across the world, including the United States [1,2]. As of May 29, 2020, there were over 1.7 million confirmed COVID-19 cases and more than 100,000 deaths in the United States [3]. In the face of a rapidly growing pandemic such as COVID-19, prompt and up-to-date assessment of the public's behavioral responses to the pandemic is critical if the findings are to be informative to public health policies and responses at local, regional, and national levels [4]. Compared with previous pandemics, the COVID-19 pandemic poses unprecedented challenges to public health responses due in part to its unique epidemiological characteristics. For example, the incubation period of COVID-19 can be as long as 24 days, and studies have found that a significant proportion of individuals infected with COVID-19 were asymptomatic but highly contagious, thus, posing enormous challenges for containing the spread of COVID-19 [5,6]. For the time being, there are no vaccines or antiviral medicine to treat or prevent this novel coronavirus [7]. Given all these factors, it is imperative for the public to actively engage in preventive behaviors and testing for the virus [8]. Correspondingly, research that identifies potential predictors of engagement in preventive behaviors and testing will generate urgently needed insights into social responses to the pandemic and inform targeted interventions to promote preventive behaviors and testing.

At the time of writing, we are aware of only one study that examined engagement in preventive measures (eg, avoiding in-person social interactions, staying home, washing hands) during the early stage (early March 2020) of the pandemic in the United States [9]. Although another study examined the use of masks as a behavioral response during the pandemic, this behavior was not examined as a preventive measure but instead as a behavioral response against the Centers for Disease Control and Prevention (CDC) and National Institute of Health (NIH) recommendations [10]. To our knowledge, no research has examined testing behaviors during the pandemic. Large-scale testing, followed by contact tracing and isolation of those with positive test results, is an essential measure for preventing a large fraction of possible transmission chains [8,11,12].

As a new infectious disease, COVID-19 has triggered a massive spike in uncertainty among the public. To learn more about the disease and to better cope with the pandemic, people are motivated to acquire relevant information through various sources. The internet has become a particularly important source of health information [13]. Recent research shows that people rely heavily on the internet to search for relevant COVID-19 health information [14]. Besides acquiring information through active search, people are also incidentally exposed to health information online [15]. Health information received online not only fills an information gap but also influences people's health decision making [16,17]. Besides the internet, personal experiences serve as a prominent means to acquire information. Because the virus is primarily transmitted through personal contacts, awareness of infection in one's social surroundings, including immediate family, friends and relatives, and local communities, are likely to affect people's risk perceptions and their engagement in preventive behaviors and testing. Past research has also shown that different demographic characteristics tend to be correlated with preventive behaviors during a pandemic [18]. A newly published article noted that demographic characteristics such as age and sex were associated with COVID-19 fatality rates [19], suggesting a necessary role of demographic factors in the investigation of behavioral responses during the COVID-19 pandemic. In this study, we examine internet use, risk awareness, and demographic characteristics associated with preventive behaviors and testing during the COVID-19 pandemic in the United States. Internet use is primarily examined via the amount of COVID-19–related information one has received online, including information received from both active search and passive exposure. Risk awareness is conceptualized as the extent that people have knowledge of infections in their social surroundings. Major demographic characteristics including sex, age, ethnicity, income, education, marital status, and employment status were also examined in this study.

Methods

Sampling Participants

This study received ethical approval from the corresponding author's university. Participants for the study were recruited from Amazon Mechanical Turk (MTurk), an online crowdsourcing labor marketplace operated by Amazon. There is evidence showing that MTurk samples provide data equivalent in quality to the data generated from alternative samples [20]. The survey was constructed and administered using Qualtrics (version 12; Qualtrics International Inc). Qualtrics records individual responses to the survey but not the MTurk account information, so participants remain anonymous. Each participant received US $0.75 for their participation.

Data Collection

Data collection started on April 10, 2020, and was completed on April 14, 2020. Upon consent, participants were instructed to complete a survey asking about their perceptions and behaviors related to the COVID-19 pandemic. Specifically, each participant was asked about whether or not they got tested for the virus, their engagement with different types of preventive behaviors, the extent to which they received information related to COVID-19, the amount of time they spent on the internet on a daily basis, and risk awareness regarding others' COVID-19–related health status (see Textbox 1). For engagement in preventative behaviors against COVID-19, items were measured on a 5-point scale (1=never, 2=sometimes, 3=about half the time, 4=most of the time, 5=always) and were prefaced with the question "Over the past month, how often have you engaged in the following practices to minimize the risk of contracting the coronavirus (COVID-19)?" For receiving COVID-19–related health information online, items were measured on a 5-point scale (1=didn't receive at all, 2=received rarely, 3=received occasionally, 4=received regularly, 5=received a great deal) and were prefaced with the question "Over the past month, to what extent have you received the following types of informational support online?” Extent of receiving COVID-19–related information was calculated based on the average of the four items (Cronbach alpha=.84). Based on the recommendations from the World Health Organization and the CDC in the United States [7,21], seven types of preventive behaviors were examined in this study: wearing a facemask, washing hands, covering nose and mouth when sneezing and coughing, social distancing, staying home, avoiding public transportation, and cleaning frequently touched surfaces. Risk awareness of others’ COVID-19–related health status was assessed by asking if the participants were aware of any positive cases in their immediate family, among close friends and relatives, or in local communities. The survey also obtained participants’ basic demographic information including sex, age, ethnicity, income, marital status, educational level, and employment status.

In total, 1080 MTurk workers filled out the online questionnaire. To ensure data quality, we included attention checks in the questionnaire. At three different places in the questionnaire, participants were asked to select a designated answer without giving a specific content question (eg, “Please select ‘Never’ for this question,” “Please select ‘Disagree’ for this question”). Failure to select the designated answer for any of the three questions was considered an indication of random clicking. This resulted in the exclusion of 101 participants. The final data set included 979 participants’ survey responses.

Textbox 1. Measurement of engagement in preventive behaviors and online information reception.

Engagement in preventive behaviors against the coronavirus disease (COVID-19)

- Wear a facemask in public even if I am not sick
- Wash hands regularly for 20 seconds, with soap and water or alcohol-based hand rub
- Cover nose and mouth with a disposable tissue or flexed elbow when coughing or sneezing
- Keep safe social distance with others
- Stay home
- Avoid using public transportation
- Clean and disinfect frequently touched surfaces such as doorknobs, phones, and keyboards daily

Received COVID-19–related health information online

- Information regarding the scientific facts (eg, symptoms, causes of the disease) related to the pandemic
- Information regarding how to prevent contracting the virus
- Information regarding the spreading of the virus
- Information regarding the sources and resources to give and receive social support during the pandemic

Data Analysis

To assess the effects of internet use, risk awareness, and demographic characteristics on engagement in preventive behaviors and testing of COVID-19, we conducted ordinal logistic regression for the 5-level self-reported engagement in preventive behaviors and binomial logistic regression analysis for the binary testing behavior. In both ordinal and binomial logistic regression analyses, participants’ *demographic characteristics* (ie, sex, age, ethnicity, income, education level, marital status, employment status) were entered in step 1 (see Model 1s in Multimedia Appendix 1), *internet use* (time spent on the internet and the extent of receiving COVID-19–related information online) was entered in step 2 (see Model 2s in Multimedia Appendix 1), and *risk awareness* (whether any immediate family members, close friends or relatives, and people in the local community tested positive for COVID-19) was entered in step 3 (see Model 3s in Multimedia Appendix 1). Detailed results are reported in Multimedia Appendix 1. SPSS 26 (IBM Corp) was used to perform the analyses.

Reported in the Results section, the findings pertaining to the effects of each demographic characteristic reflect their independent effects controlling for other demographic characteristics (ie, based on Model 1s in Multimedia Appendix 1). Findings pertaining to the effects of internet use or risk awareness reflect their independent effects controlling for all other predictors (ie, based on Model 3s in the Multimedia Appendix 1).

Results

Participant Characteristics

Demographic characteristics of the 979 participants are provided in Table 1.

Table 1. Sample characteristics (n=979).

Demographics	Participants
Female, n (%)	466 (47.6)
Age groups (years), n (%)	
18-27	210 (21.5%)
28-37	391 (39.9%)
38-47	196 (20.0%)
48-57	104 (10.6%)
≥58	78 (8%)
Age (years), mean (SD)	36.94 (11.93)
Marital status, n (%)	
Single	**449 (45.9)**
Single	376 (38.4)
Widowed	8 (0.8)
Divorced	54 (5.5)
Separated	11 (1.1)
Married/domestic partnership	530 (54.1)
Employment status, n (%)	
Working	**793 (81.0)**
Self-employed	134 (13.7)
Working full time for wages	568 (58)
Working part time for wages	91 (9.3)
Not working	**186 (19.0)**
Out of work	97 (9.9)
Not able to work or disabled	18 (1.8)
Retired	25 (2.6)
Other	46 (4.7)
Race, n (%)	
White	677 (69.2)
Black or African American	112 (11.4)
Hispanic or Latino American	46 (4.7)
Asian or Asian American	96 (9.8)
Other	48 (4.9)
Education level, n (%)	
Less than high school degree	4 (0.4)
High school graduate	63 (6.4)
Some college but no degree	148 (15.1)
Associate degree in college	104 (10.6)
Bachelor’s degree in college	447 (45.7)
Master’s degree	176 (18.0)
Doctoral degree	21 (2.1)
Professional degree (JD, MD)	16 (1.6)
Education year[a], mean (SD)	15.62 (2.15)
Household income (US $), n (%)	

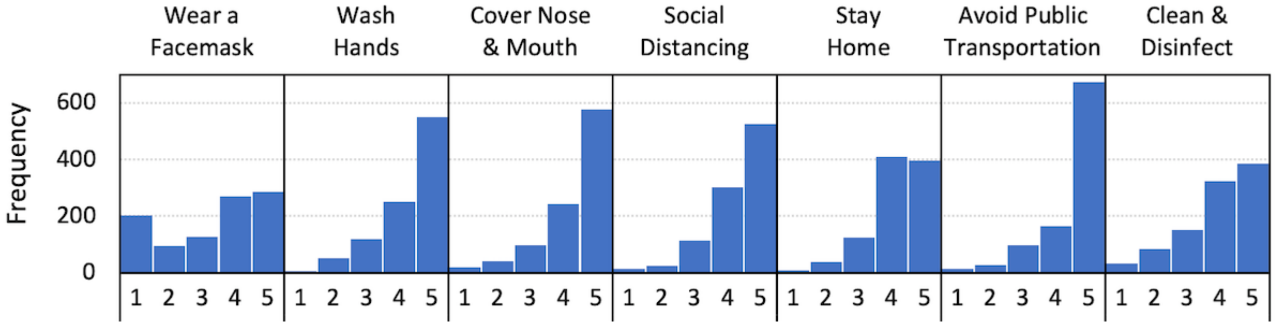
Demographics	Participants
<10,000	45 (4.6)
10,001-20,000	76 (7.8)
20,001-40,000	183 (18.7)
40,001-60,000	222 (22.7)
60,001-80,000	201 (20.5)
80,001-100,000	106 (10.8)
100,001-120,000	64 (6.5)
>120,000	82 (8.4)

[a] Education year is transformed from education level based on typical years of completion.

Engagement in Preventive Behaviors

As shown in Figure 1, participants reported less frequent engagement in wearing a facemask in public than other preventive measures (mean 3.35, SD 1.50), and a fair number of participants reported that they never wore a facemask during the pandemic. A large number of participants reported that they frequently washed hands (mean 4.31, SD 0.93), covered nose and mouth when sneezing and coughing (mean 4.35, SD 0.95), kept social distance (mean 4.33, SD 0.87), stayed home (mean 4.17, SD 0.86), avoided using public transportation (mean 4.49, SD 0.89), and cleaned frequently touched surfaces (mean 3.97, SD 1.09) to protect themselves from COVID-19.

Figure 1. Response distributions of the self-reported preventative behaviors. Responses were coded from 1 (Never) to 5 (Always).



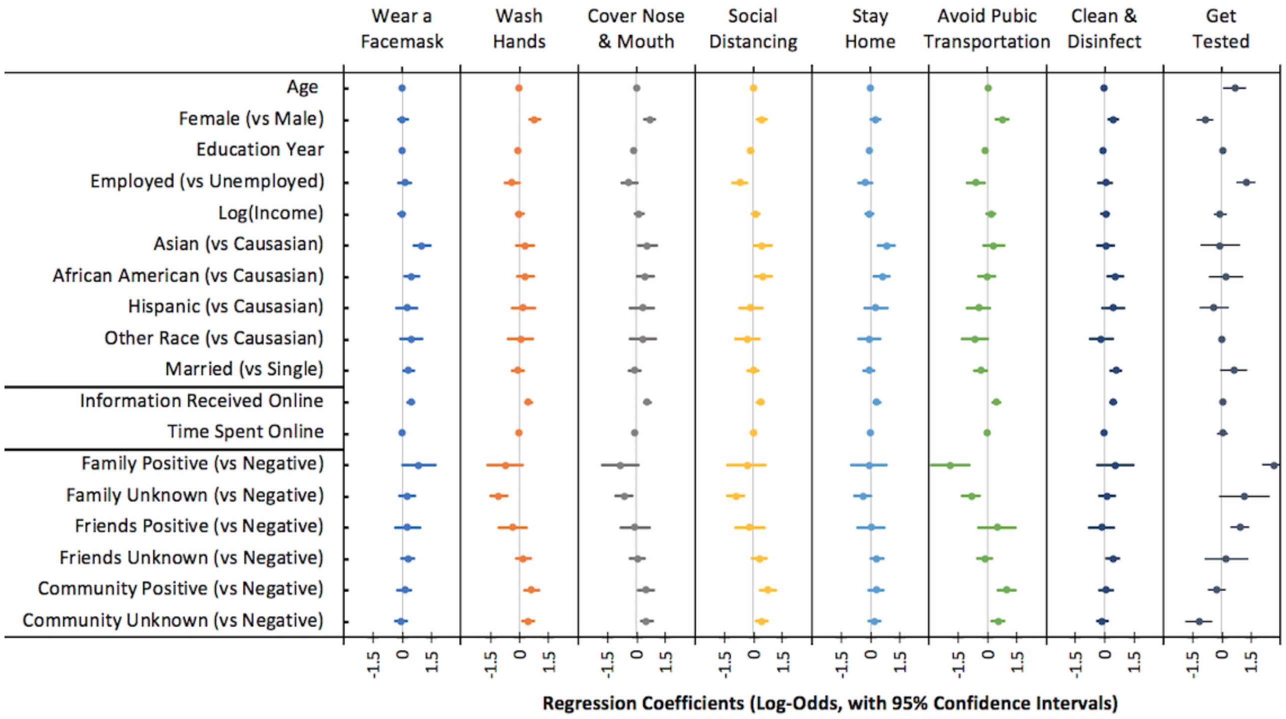
Internet Use and Engagement in Preventive Behaviors

Our data showed that the amount of COVID-19–related health information received online was positively associated with engagement in all types of preventive behaviors: wearing a facemask in public (odds ratio [OR] 1.55, 95% CI 1.34-1.79, *P*<.001), washing hands (OR 1.58, 95% CI 1.35-1.85, *P*<.001), covering nose and mouth when sneezing and coughing (OR 1.78, 95% CI 1.52-2.10, *P*<.001), keeping social distance with others (OR 1.41, 95% CI 1.21-1.65, *P*<.001), staying home (OR 1.40, 95% CI 1.20-1.62, *P*<.001), avoiding using public transportation (OR 1.57, 95% CI 1.32-1.88, *P*<.001), and cleaning frequently used surfaces (OR 1.55, 95% CI 1.34-1.79, *P*<.001).

Risk Awareness and Engagement in Preventive Behaviors

Our data revealed that awareness of immediate family members' test results was associated with participants' engagement in preventive behaviors (see Figure 2 and Model 3s in Multimedia Appendix 1). Specifically, compared with participants who did not have immediate family members with positive test results, those who had immediate family members with positive results less often washed hands (OR 0.35, 95% CI 0.23-0.52, *P*<.001), covered nose and mouth when sneezing and coughing (OR 0.53, 95% CI 0.35-0.80, *P*=.003), kept social distance with others (OR 0.40, 95% CI 0.26-0.61, *P*<.001), and avoided using public transportation (OR 0.42, 95% CI 0.28-0.65, *P*<.001).

Figure 2. Summaries of effect estimates (standardized) and 95% CIs based on logistic regression models predicting self-reported preventative behaviors (ordered log-odds) and testing behaviors (log-odds).



Participants reported more frequent effort in cleaning frequently touched surfaces if they had close friends or relatives who tested positive (OR 1.52, 95% CI 1.10-2.10, *P*=.01). In addition, participants who were aware of positive cases in their local communities reported more frequent hand washing (OR 1.54, 95% CI 1.15-2.06, *P*=.004), covering nose and mouth when sneezing and coughing (OR 1.68, 95% CI 1.25-2.26, *P*=.005), keeping social distance (OR 1.51, 95% CI 1.13-2.02, *P*<.001), and avoiding using public transportation (OR 1.71, 95% CI 1.24-2.37, *P*=.001) than those who reported no positive cases in their communities.

Demographic Characteristics and Engagement in Preventive Behaviors

As shown in Figures 2 and 3, several demographic characteristics were found to be associated with engagement in preventive behaviors (also see Model 1s in Multimedia Appendix 1). Compared with males, females more frequently washed their hands (OR 2.39, 95% CI 1.85-3.09, *P*<.001), covered their nose and mouth when sneezing and coughing (OR 2.12, 95% CI 1.63-2.74, *P*<.001), kept social distance with others (OR 1.64, 95% CI 1.28-2.11, *P*<.001), stayed home (OR 1.34, 95% CI 1.05-1.70, *P*=.02), avoided using public transportation (OR 2.30, 95% CI 1.72-3.07, *P*<.001), and cleaned frequently touched surfaces (OR 1.58, 95% CI 1.25-2.00, *P*<.001).

Compared with younger participants, older participants reported more frequent efforts to wash hands (OR 1.01, 95% CI 1.00-1.02, *P*=.04), cover nose and mouth when sneezing and coughing (OR 1.01, 95% CI 1.00-1.03, *P*=.02), keep social distance (OR 1.02, 95% CI 1.01-1.03, *P*=.001), stay home (OR 1.01, 95% CI 1.00-1.02, *P*=.02), and avoid public transportation (OR 1.02, 95% CI 1.01-1.04, *P*<.001) than younger participants.

Ethnic differences were also observed in engagement with preventive behaviors. Compared with whites, African Americans and Asians more frequently wore a facemask in public (OR 1.81, 95% CI 1.26-2.59, *P*<.001; OR 2.47, 95% CI 1.65-3.69, *P*<.001, respectively) and stayed home (OR 1.88, 95% CI 1.28-2.77, *P*=.001; OR 2.23, 95% CI 1.47-3.37, *P*<.001, respectively). In addition, compared with whites, Asians covered their noses and mouths when sneezing and coughing more often (OR 1.78, 95% CI 1.13-2.80, *P*=.01), and kept social distance more often (OR 1.61, 95% CI 1.04-2.48, *P*=.03). African Americans reported more frequent effort in cleaning frequently touched surfaces than whites (OR 2.00, 95% CI 1.36-2.94, *P*<.001).

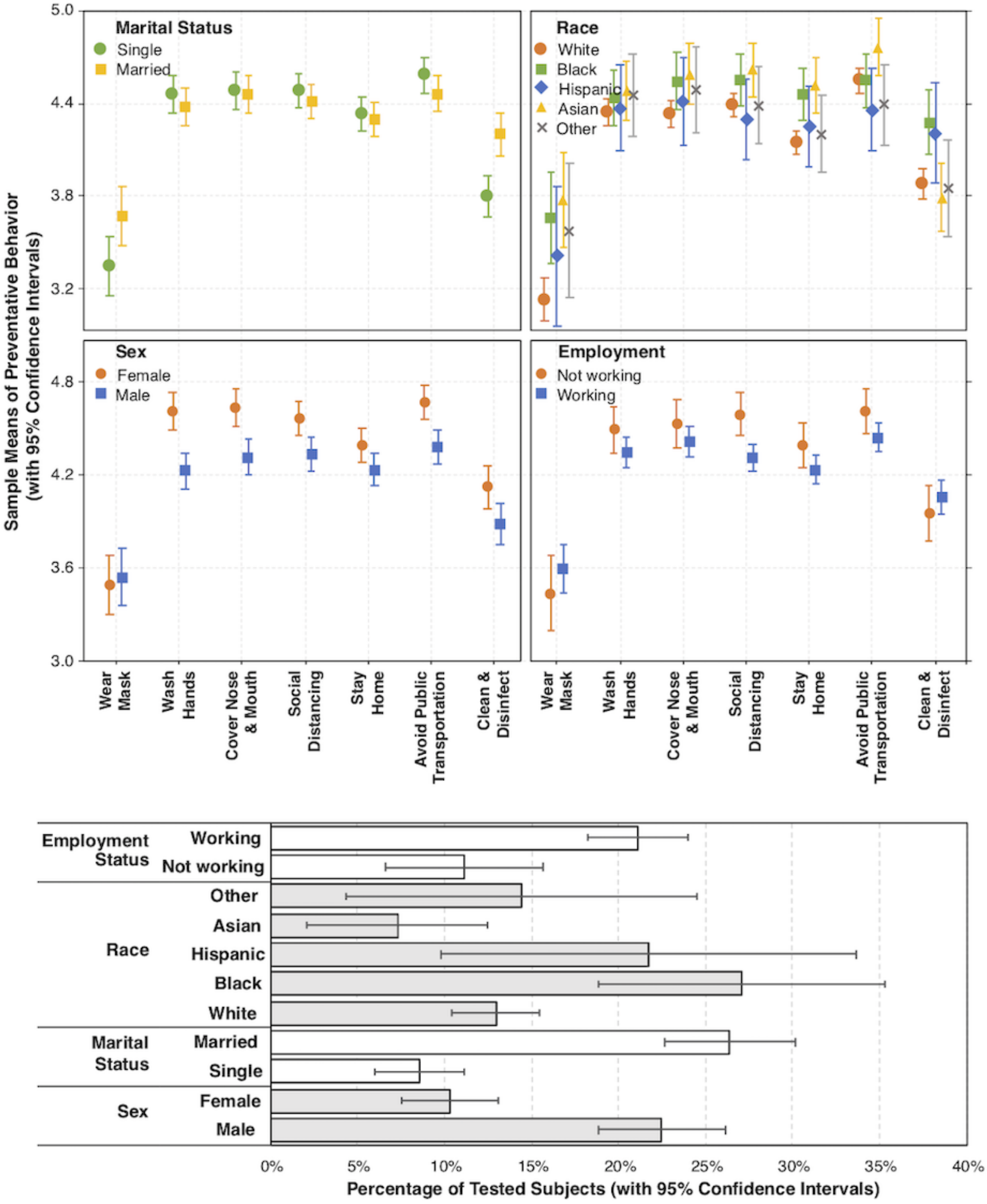
More educated participants less frequently engaged in the following preventive behaviors: washing hands (OR 0.93, 95% CI 0.87-0.99, *P*=.03), covering nose and mouth when sneezing and coughing (OR 0.87, 95% CI 0.81-0.93, *P*<.001), keeping social distance (OR 0.90, 95% CI 0.84-0.96, *P*=.001), avoiding using public transportation (OR 0.88, 95% CI 0.82-0.95, *P*<.001), and cleaning and disinfecting touched surfaces (OR 1.25, 95% CI 1.04-1.50, *P*=.02).

Participants who are married or have domestic partners more frequently wore a facemask in public (OR 1.43, 95% CI 1.11-1.84, *P*=.01), kept social distance with others (OR 1.64, 95% CI 1.28-2.11, *P*<.001), and cleaned frequently touched surfaces (OR 1.84, 95% CI 1.42-2.38, *P*<.001) than their single counterparts. However, compared to single participants, participants who are married or have domestic partners less frequently avoided public transportation (OR 0.90, 95% CI 0.84-0.96, *P*=.001).

Compared with unemployed participants, employed participants less frequently kept social distance (OR 0.50, 95% CI 0.35-0.72, *P*<.001) or avoided using public transportation (OR 0.56, 95% CI 0.36-0.86, *P*=.01).

Compared to the participants who reported having lower income, the ones reported to have higher income covered their noses and mouths more often (OR 1.25, 95% CI 1.04-1.50, *P*=.02), and avoided public transportation more often (OR 1.29, 95% CI 1.06-1.58, *P*=.01).

Figure 3. Sample means of self-reported preventative behaviors and proportions of testing (with 95% CIs) across categorical demographic characteristics.



Testing Behavior

Out of the 979 participants, 22.7% (n=222/979) reported that they had tested for COVID-19, including 23.0% (n=51/222) who reported positive results, 72.5% (n=161/222) who reported negative results, and 4.5% (n=10/222) who did not know the results at the time of participation. Of the 222 participants who got tested, 44.6% (n=99) had at least one positive case in their immediate family, 45.5% (n=101) had at least one positive case among close friends or relatives, and 51.3% (n=114) had at least one positive case in their local communities.

The extent of receiving COVID-19–related health information online was not associated with their odds of testing (OR 1.03, 95% CI 0.81-1.32, *P*=.79). Awareness of COVID-19 infection in one's social surroundings was associated with odds of testing, although the patterns slightly differed between close and distant social circles. For participants who had positive cases in their immediate family, their odds of testing were much higher than those who did not have any positive cases in their immediate family (OR 1.48, 95% CI 8.28-26.44, *P*<.001). Similarly, for participants who had any close friends or relatives that tested positive, their odds of testing were much higher compared with those who had no close friends or relatives with positive test results (OR 2.52, 95% CI 1.58-4.03, *P*<.001). Participants who were aware of positive cases in their local communities did not appear to differ in testing from those who were not aware or uncertain if there were any positive cases in local communities.

Results of the logistic regression showed that several demographic characteristics were associated with testing for COVID-19. Specifically, the odds of females getting tested were only 39.75% (95% CI 28.12%-56.20%) of those of males (*P*<.001). The odds of African Americans getting tested were 249.98% (95% CI 156.95%-398.14%) of those of white Americans (*P*<.001) and 275.45% (95% CI 118.73%-639.01%) of those of Asian Americans (*P*=.02). Older participants' odds of testing were 96.39% (95% CI 94.76%-98.05%) of those who were 1 year younger (*P*<.001). The odds of testing among participants who are married or have domestic partners were 383.92% (95% CI 260.70%-565.36%) of those of singles (*P*<.001).

Discussion

Principal Findings

This study examined if internet use, risk awareness, and demographic characteristics were associated with engagement in preventive behaviors and testing for COVID-19 in the United States. Taken as a whole, our data revealed several notable patterns of findings pertaining to preventive behaviors.

First, consistent with findings from other studies on individuals' engagement in preventive measures against COVID-19 [9,22,23], our data showed that there is a greater tendency for people to engage in some preventative behaviors than others. Although there is evidence showing that, as the pandemic progresses, individuals develop greater awareness of the health risk posed by the virus and engage in protective behaviors with increasing frequency [9], we found that this development is unequal across different types of preventive behaviors. Specifically, people seemed to be more active in adopting the preventive measures of washing hands, covering mouths when sneezing, keeping social distance, and avoiding public transportation than wearing a facemask, staying home, or cleaning frequently touched surfaces. This finding is consistent with several recent studies that assessed the public's perceptions or knowledge about COVID-19 [9]. In particular, people have different perceptions of the effectiveness or necessity of various preventive measures. For example, although wearing a facemask in public has been a mandatory preventive measure in some countries [23] and many states in the United States, its utility as a preventive measure against the coronavirus is still highly controversial [24]. In addition, as some scholars have noted, the practice of mask wearing is an evolving and cultural phenomenon [10]. Although older people in the United States were found to be less likely to wear a mask during the early stage of the pandemic, which is likely due to their higher knowledge of CDC and NIH recommendations against mask use [10], our survey showed that they seemed to have quickly adapted their behaviors following nationwide revised recommendations for mask use. At the same time, different levels of engagement in preventive behaviors during the COVID-19 pandemic can be attributed to the fact that some preventive measures require greater effort (eg, cleaning frequently touched surfaces) or pose greater difficulties (eg, staying home) than others and are thus deemed less feasible.

Second, we observed positive associations between the extent of receiving COVID-19–related information online and engagement in all types of preventive behaviors. It is possible that a good proportion of the COVID-19–related information people received or sought online involved recommendations on preventive measures [25], leading to more frequent engagement in preventive behaviors. At the same time, this finding suggests that receiving pandemic-related information online, despite vast variation in information content, may enhance people's concerns about the pandemic and motivate them to actively take preventive measures [26].

Third, risk awareness regarding infection in one's immediate family, close friends and relatives, and local communities was differentially associated with engagement in preventative behaviors. Compared with participants who did not have positive cases in their immediate family, those who had positive cases in their immediate family reported less frequent engagement in almost all preventive behaviors. Although this finding is somewhat counterintuitive, it also implies that a lack of prevention may lead to higher risk of infection in a family. Immediate family members, as the closest contacts, are at high risk of transmitting the virus within the household. This finding suggests that taking preventive behaviors not only helps one protect themself but also reduces risks of their immediate families becoming infected. Awareness of positive cases among friends and relatives seemed to have limited influence on one's preventative behaviors. A possible explanation is that many people do not live in close proximity with their friends and relatives, and thus do not perceive high risk of infection from them. One's local communities appeared to have a positive impact on engagement with preventive behaviors. Compared with participants who reported no positive cases in their communities, those who were aware of positive cases in their communities more frequently washed hands, covered nose and mouth when coughing, kept social distance, and minimized using public transportation. This finding suggests that one's behavioral responses to a pandemic is influenced by immediate risk factors in one's surroundings.

Fourth, subgroups tend to differ in their engagement in preventive behaviors. For example, consistent with many past studies, our data showed that being older, female [27-32], or nonwhite [33] is associated with a higher chance of adopting preventive behaviors during a pandemic involving respiratory type diseases. This finding is also consistent with recent research on COVID-19 showing that older people and females were more knowledgeable about COVID-19 [10].

In addition, working people reported less frequent engagement in terms of avoiding public transportation and keeping social distance. This finding suggests that working individuals, especially those who cannot work from home, face greater challenges in implementing certain preventative measures. In particular, for individuals who work in essential businesses (eg, supermarkets, health care, post office, food processing factories), it may not be feasible for them to employ preventive measures such as keeping social distance or avoiding public transportation.

Our findings pertaining to testing behaviors seemed less straightforward. The extent of receiving health information online was not associated with testing. This may be largely due to the fact that during the early stage of the pandemic eligibility criteria for testing were highly stringent, and testing capacity was limited in most parts of the country. Although individuals can request a test for COVID-19, whether or not one will get tested is a decision ultimately made by health departments and health professionals. In other words, unlike many preventive behaviors that can be performed based on one's own volition, individuals have much less power in decision making regarding testing.

Not surprisingly, individuals who had confirmed COVID-19 cases in their immediate family were more likely to get tested. Individuals who had confirmed COVID-19 cases among their close friends and relatives were also likely to get tested. On the other hand, positive cases in one's local communities did not seem to have a significant influence on one's testing. These findings are consistent with CDC's guidelines for testing, one of which is close contact with patients who are infected. As testing capacity increases and testing criteria become less stringent, it is reasonable to expect that awareness of positive cases in local communities would motivate more individuals to request testing.

Our study also reveals some subgroup differences in testing. For example, we found that working individuals were twice as likely as nonworking individuals to get tested for COVID-19. In addition, individuals who are single were much less likely than those who are married or have domestic partners to get tested. Given that individuals may contract the coronavirus at virtually any setting with others around (eg, grocery stores, parks, social gatherings), this finding suggests that some nonworking individuals or those who live by themselves might have underestimated their risk of COVID-19 infection.

Our findings offer several implications for interventions, communication strategies, and future research. First, in light of recent research showing that many health care workers had poor knowledge of the mode of transmission and the incubation period of COVID-19 [33], both of which can affect health care workers' recommendations to their patients regarding preventive behaviors and testing (eg, asymptomatic patients often do not get tested), there is an urgent need to provide health care workers with up-to-date information about the disease. Second, as shelter-in-place orders remain in effect in most states in the United States, greater efforts should be put into increasing household internet coverage so that more people will have easy and prompt access to information related to the pandemic [34]. Third, our findings suggest a need for more public health education programs and interventions targeting certain subgroups that have consistently shown to be less likely to adopt preventive measures during a pandemic. Identifying more effective strategies that can be used to induce self-protective behaviors in groups such as young males, for example, can help slow down the spread of the virus. At the same time, as other scholars have noted, although demographic characteristics are generally immutable, future research needs to obtain a deeper understanding of the root causes of differential behavioral responses, which can help inform the development of dissemination strategies directed at different subgroups [28].

Limitations

This study has several limitations. First, given the cross-sectional nature of our data, we are not in a position to empirically assess and demonstrate causal relationships among the variables. For instance, confirmed cases in one's social surroundings may prompt people to engage in more active preventive behaviors, and one's engagement in preventive behaviors during a pandemic like COVID-19 can certainly influence the health of families, friends, and people in local communities.

Second, given the urgency of the research needs and limited access to nationally representative samples, we elected to recruit our participants from an online crowdsourcing platform. Although the composition of ethnicity, sex, and marital status in our data largely mirrored the demographics profile in the US population [35-37], our sample included more young, educated, and working participants. Ideally, a randomly selected sample of the public should be surveyed and comparisons made with the known distribution of key variables in the population.

Third, although social desirability bias has been found to be lower in anonymous online surveys than in telephone or in-person surveys [38], we cannot rule out the possibility of some response bias in the self-reported data. In particular, potential group variations in response bias may have influenced some findings. For example, past research suggests that females tend to show more social desirability than males in survey responses [39]. Although the observed pattern of gender differences in this study is largely consistent with past research, we still cannot rule out the possibility that female participants reported more frequent engagement in preventive behaviors than male participants due, in part, to social desirability. Despite this limitation, a recent study [4] has validated the use of rapid surveys to examine COVID-19–related perceptions and behaviors.

Conclusions

During the ongoing and rapidly evolving COVID-19 pandemic, we have seen a bourgeoning amount of research pertaining to the pathophysiology, diagnosis, and treatment of the novel disease, which is urgently and rightfully needed. Research examining the public's behavioral (and cognitive, psychological) responses to the pandemic, however, also deserves attention, as it can help inform formulation and implementation of public health policies and control measures. This study offers useful insights into factors that are associated with engagement in preventive behaviors and testing of COVID-19. Our findings revealed that the extent of receiving COVID-19–related information online, risk awareness, and demographic characteristics including sex, ethnicity, age, marital status, and employment status are key factors associated with individuals' engagement in various preventive behaviors and testing for COVID-19.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Binomial and ordinal logistic regression results on testing and engagement in preventive behaviors.
[DOCX File , 77 KB - jmir_v22i6e19782_app1.docx]

Multimedia Appendix 2
Simplified questionnaire.
[DOCX File , 28 KB - jmir_v22i6e19782_app2.docx]

References

1. Fernandes N. Economic effects of coronavirus outbreak (COVID-19) on the world economy. SSRN J 2020 Mar 24:24. [doi: 10.2139/ssrn.3557504]
2. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? Lancet 2020 Apr;395(10231):1225-1228. [doi: 10.1016/s0140-6736(20)30627-9]
3. Centers for Disease Control and Prevention. Cases in U.S URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html [accessed 2020-05-29]
4. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: a cross-sectional urvey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
5. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro Surveill 2020 Jan;25(4):1-5 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.4.2000058] [Medline: 32019669]
6. Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. J Med Virol 2020 Mar 05:568-576 [FREE Full text] [doi: 10.1002/jmv.25748] [Medline: 32134116]
7. World Health Organization. 2020 Apr 17. Q&A on coronaviruses (COVID-19) URL: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses [accessed 2020-04-27]
8. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. Swiss Med Wkly 2020 Mar 09;150:w20225. [doi: 10.4414/smw.2020.20225] [Medline: 32191813]
9. Wise T, Zbozinek T, Michelini G, Hagan CC, Mobbs D. Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States. PsyArXiv 2020 Mar 19:1-13. [doi: 10.31234/osf.io/dz428]
10. Clements JM. Knowledge and behaviors toward COVID-19 among US residents during the early days of the pandemic: cross-sectional online questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]
11. Sharfstein JM, Becker SJ, Mello MM. Diagnostic testing for the novel coronavirus. JAMA 2020 Mar 09:1437-1438. [doi: 10.1001/jama.2020.3864] [Medline: 32150622]
12. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Global Health 2020 Apr;8(4):e488-e496. [doi: 10.1016/s2214-109x(20)30074-7]
13. Fox S, Duggan M. Pew Research Center. 2013 Jan 15. Health online 2013 URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-04-27]
14. Cuan-Baltazar J, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
15. Tian Y, Robinson JD. Incidental health information use on the internet. Health Commun 2009 Jan;24(1):41-49. [doi: 10.1080/10410230802606984] [Medline: 19204857]
16. Iverson S, Howard K, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]

17. Li S, Hubner A. The impact of web-based ratings on patient choice of a primary care physician versus a specialist: randomized controlled experiment. J Med Internet Res 2019 Jun 28;21(6):e11188 [FREE Full text] [doi: 10.2196/11188] [Medline: 31254337]
18. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]
19. Dowd J, Andriano L, Brazel D, Rotondi V, Block P, Ding X, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proc Natl Acad Sci U S A 2020 May 05;117(18):9696-9698 [FREE Full text] [doi: 10.1073/pnas.2004911117] [Medline: 32300018]
20. Sheehan K. Crowdsourcing research: data collection with Amazon’s Mechanical Turk. Commun Monogr 2017 Jul 04;85(1):140-156. [doi: 10.1080/03637751.2017.1342043]
21. Centers for Disease Control and Prevention. How to protect yourself & others URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html [accessed 2020-04-30]
22. Villarreal-Benavides ME. COVID-19 in Mexico: Perception and knowledge of the pandemic in the Mexican population through the application of online surveys. JMIR Preprints 2020 Apr 27:1-27. [doi: 10.2196/preprints.19654]
23. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
24. Feng S, Shen C, Xia N, Song W, Fan M, Cowling B. Rational use of face masks in the COVID-19 pandemic. Lancet Respir Med 2020 May;8(5):434-436 [FREE Full text] [doi: 10.1016/s2213-2600(20)30134-x]
25. Kwok KO, Li KK, Chan HH, Yi YY, Tang A, Wei WI, et al. Community responses during the early phase of the COVID-19 epidemic in Hong Kong: risk perception, information exposure and preventive measures. medRxiv 2020 Feb 27:1-26. [doi: 10.1101/2020.02.26.20028217]
26. Kim Y, Zhong W, Jehn M, Walsh L. Public risk perceptions and preventive behaviors during the 2009 H1N1 influenza pandemic. Disaster Med Public Health Prep 2015 Apr;9(2):145-154. [doi: 10.1017/dmp.2014.87] [Medline: 25882121]
27. Lau JTF, Yang X, Tsui H, Kim JH. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10 to day 62. J Epidemiol Community Health 2003 Nov;57(11):864-870 [FREE Full text] [doi: 10.1136/jech.57.11.864] [Medline: 14600111]
28. Leung G, Ho L, Chan S, Ho S, Bacon-Shone J, Choy R, et al. Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong. Clin Infect Dis 2005 Jun 15;40(12):1713-1720. [doi: 10.1086/429923] [Medline: 15909256]
29. Quah SR, Hin-Peng L. Crisis prevention and management during SARS outbreak, Singapore. Emerg Infect Dis 2004 Feb;10(2):364-368 [FREE Full text] [doi: 10.3201/eid1002.030418] [Medline: 15030714]
30. Tang CS, Wong C. Factors influencing the wearing of facemasks to prevent the severe acute respiratory syndrome among adult Chinese in Hong Kong. Prev Med 2004 Dec;39(6):1187-1193 [FREE Full text] [doi: 10.1016/j.ypmed.2004.04.032] [Medline: 15539054]
31. Jones JH, Salathé M. Early assessment of anxiety and behavioral response to novel swine-origin influenza A(H1N1). PLoS One 2009 Dec 03;4(12):e8032 [FREE Full text] [doi: 10.1371/journal.pone.0008032] [Medline: 19997505]
32. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul 02;339:b2651 [FREE Full text] [doi: 10.1136/bmj.b2651] [Medline: 19574308]
33. Bhagavathula AS, Aldhaleei WA, Rahmani J, Mahabadi MA, Bandari DK. Knowledge and perceptions of COVID-19 among health care workers: cross-sectional study. JMIR Public Health Surveill 2020 Apr 30;6(2):e19160 [FREE Full text] [doi: 10.2196/19160] [Medline: 32320381]
34. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 2020 Feb 25;12(3):254 [FREE Full text] [doi: 10.3390/v12030254] [Medline: 32106567]
35. United States Census Bureau. 2020 Feb 05. Marital status in United States URL: https://www.census.gov/library/visualizations/interactive/marital-status-in-united-states.html [accessed 2020-04-30]
36. U.S. Bureau of Labor Statistics. Labor force statistics from the current population survey URL: https://www.bls.gov/cps/cpsaat08.htm [accessed 2020-04-30]
37. United States Census Bureau. QuickFacts URL: https://www.census.gov/quickfacts/fact/table/US/PST045219 [accessed 2020-04-30]
38. Keeter S. Pew Research Center. 2015 May 13. From telephone to the web: the challenge of mode of interview effects in public opinion polls URL: https://www.pewresearch.org/methods/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/
39. Phillips DL, Clancy KJ. Some effects of "social desirability" in survey studies. Am J Sociol 1972 Mar;77(5):921-940. [doi: 10.1086/225231]

Abbreviations

CDC: Centers for Disease Control and Prevention
COVID-19: coronavirus disease
MTurk: Amazon Mechanical Turk
NIH: National Institute of Health
OR: odds ratio

Edited by G Eysenbach; submitted 01.05.20; peer-reviewed by OS Liang, L Sbaffi, R El Sherif; comments to author 23.05.20; revised version received 30.05.20; accepted 04.06.20; published 16.06.20.

Please cite as:
Li S, Feng B, Liao W, Pan W
Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States
J Med Internet Res 2020;22(6):e19782
URL: http://www.jmir.org/2020/6/e19782/
doi:10.2196/19782
PMID:32501801

Original Paper

Clarification of Misleading Perceptions of COVID-19 Fatality and Testing Rates in Italy: Data Analysis

Davide Tosi[1], PhD; Alessandro Verde[2*], MD, PhD; Manuela Verde[3*], MD, MSc

[1]Department of Theoretical and Applied Science, University of Insubria, Varese, Italy

[2]Heart Failure Unit, Niguarda Cardio Center, Cardiothoracic and Vascular Department, ASST Great Metropolitan Hospital Niguarda, Milan, Italy

[3]CRIMEDIM Research Centre in Emergency and Disaster Medicine, Translational Medicine Department, Università del Piemonte Orientale, Novara, Italy

[*]these authors contributed equally

Corresponding Author:
Davide Tosi, PhD
Department of Theoretical and Applied Science
University of Insubria
Jean Henry Dunant, 3
Varese, 21100
Italy
Phone: 39 0332 42138
Email: davide.tosi@uninsubria.it

Abstract

Background: The fatality rate of coronavirus disease (COVID-19) in Italy is controversial and is greatly affecting discussion on the impact of containment measures that are straining the world’s social and economic fabric, such as instigating large-scale isolation and quarantine, closing borders, imposing limits on public gatherings, and implementing nationwide lockdowns.

Objective: The scientific community, citizens, politicians, and mass media are expressing concerns regarding data suggesting that the number of COVID-19–related deaths in Italy is significantly higher than in the rest of the world. Moreover, Italian citizens have misleading perceptions related to the number of swab tests that have actually been performed. Citizens and mass media are denouncing the coverage of COVID-19 swab testing in Italy, claiming that it is not in line with that in other countries worldwide.

Methods: In this paper, we attempt to clarify the aspects of COVID-19 fatalities and testing in Italy by performing a set of statistical analyses that highlight the actual numbers in Italy and compare them with official worldwide data.

Results: The analysis clearly shows that the Italian COVID-19 fatality and mortality rates are in line with the official world scenario, as are the numbers of COVID-19 tests performed in Italy and in the Lombardy region.

Conclusions: This up-to-date analysis may elucidate the evolution of the COVID-19 pandemic in Italy.

(J Med Internet Res 2020;22(6):e19825) doi:10.2196/19825

KEYWORDS

COVID-19; SARS-CoV-2; fatality rate; swab tests; Italy; Lombardy region

Introduction

Since February 2020, the severe acute respiratory syndrom coronavirus 2 (SARS-CoV-2) pandemic has generated much concern among citizens and mass media in Italy regarding the real and official numbers provided by the Italian Government and the Italian Department of Civil Protection [1-4]. One of the most discussed topics is why the case fatality rate in Italy appears to be much higher than in other countries. Another main topic of debate is why inappropriate numbers of COVID-19 tests appear to have been performed on citizens of Italy in general and the Lombardy region in particular, thus underrepresenting the infected population.

The purpose of this study is to use a set of statistical and data analyses to clarify both the actual and official case fatality rate in Italy in comparison to those in other countries as well as the actual size of the tested population.

Methods

We examined data regarding coronavirus disease (COVID-19) case fatality rates and numbers of throat swab and nasal swab tests using real time polymerase–chain reaction assay methods conducted in world populations to describe the actual global picture of the case fatality rate of COVID-19 in Italy compared to other industrialized countries and the number of tests performed since the first outbreak in Italy (Codogno, Lodi province, and the Lombardy region) between February 24 and April 14, 2020. This time interval was selected to cover most of the COVID-19 epidemic curve. Raw data for the case fatality rate and swab tests were obtained as comma-separated value (.csv) open files from the Italian Department of Civil Protection [1] and from the Worldometer portal [5]. Data regarding the critical situation in the Lombardy region were obtained as open data directly from [1]. It is important to highlight that at the time of writing, the March 22, 2020 decree (Implementation of the decree 23 February 2020 No. 6: Urgent measures for the containment and management of the COVID-19 epidemiological emergency (20A01807), Official Gazette: General Series No.76 of 22-03-2020 [6]) is in action to temporarily stop all unnecessary industrial and commercial production activities. Currently, all of Italy is in a quarantine red zone.

To compare the situations in Italy and in other countries, the analysis was limited to the Group of Ten (G10) industrialized countries: Belgium, Canada, France, Germany, Japan, Italy, the Netherlands, the United Kingdom, the United States, and Sweden. For the Lombardy region analysis, the most affected regions and areas in each G10 country were identified and compared. The case fatality rate trend in Italy was plotted using Excel software (Microsoft Corporation), while all the other statistical analyses (absolute values, mean values, and normalized values) are reported in tables.

Results

COVID-19 Fatality Rate in Italy

To clarify both actual and official case fatality rates in comparison to those of other countries as well as the actual extent of the tested population, it is important to differentiate the case fatality rate from the actual fatality rate (infection fatality rate). The case fatality rate is the ratio of the number of deaths to the total number of positive tests, while the infection fatality rate is the ratio of the number of deaths to the total infected population. However, because the entire population (symptomatic and asymptomatic) cannot be tested with sufficient speed, it is necessary to rely on case fatality rates for epidemiology measures and policy planning. Moreover, as the epidemic escalates, the tested population represents the emergent cluster of patients who, by seeking care, turn to and contact national health systems.

At present, the COVID-19 case fatality rate in Italy is 12.80% [5]. In China, the case fatality rate shows a nonlinear trend, with a high rate at the start of the outbreak (17.3%) that decreased to 0.7% at the end of the emergency. China hit the infection curve plateau on February 20, 2020. At the time of writing, the Chinese case fatality rate is approximately 4.1% [5].

The current Italian case fatality rate is comparable with that of other European countries, such as Belgium (12.80%), the United Kingdom (12.70%), France (11.20%), and Spain (10.44%). To date, the worldwide case fatality rate is 6.40% [5]. In Germany, the case fatality rate is 2.7%; however, the Robert Koch Institute has stated that this rate is expected to grow greatly [7]. Furthermore, general misalignment with other European countries is observed for other German indices related to COVID-19, such as the distribution of deaths per age of infected people and the distribution of deaths per gender [8,9].

The Lombardy region is the region most affected by COVID-19 worldwide; data indicate that this region represents 38% of the total infections in Italy, with a regional case fatality rate of 18.08% [1]. No other countries have such unbalanced distribution and variability among their regions [5]; therefore, we can perform the statistical exercise of considering the Lombardy region value as an outlier (ie, a data point that differs significantly from the other regional observations). Hence, if we attempted to disaggregate the contribution of the Lombardy region proportionally to the national datum, we would obtain an Italian case fatality rate of 9.60%. The Italian case fatality rate value would decrease even further to 7.9% if data from the Emilia-Romagna region (the second most affected Italian region) were also excluded.

This statistical exercise stems from the consideration that the very first COVID-19 outbreak in Europe occurred in the Lombardy region, and the National Health System was realistically unprepared to face a pandemic.

Ultimately, it is important to note that the case fatality rate trend is not linear over time due to the number of total infected people depending on the daily number of new infections combined with the numbers of recovered and discharged patients. As shown in Figure 1, the trend line (ie, the 5-day moving average) of daily recoveries grows strongly over time, whereas the trend line for COVID-19 fatalities decreases with a different slope function.

Figure 2 also shows that the curve of total recovered/discharged patients grows much faster than the curve of total deaths. The difference in the slopes of the two curves could enable a reliable prediction of a decrease in the national case fatality rate at the end of the first wave of the epidemic.

Furthermore, as officially reported in the available open data [1], the occupancy of beds in COVID-19–dedicated intensive care units is steadily decreasing, with increasingly diverging curves between the number of new infections and the number of patients with severe symptoms that could lead to death in the near future. At the time of writing, there are more than 3000 patients with severe symptoms out of 104,000 total active positive cases.

Figure 1. Graphical plot of COVID-19 daily fatalities versus recoveries in Italy. COVID-19: coronavirus disease.

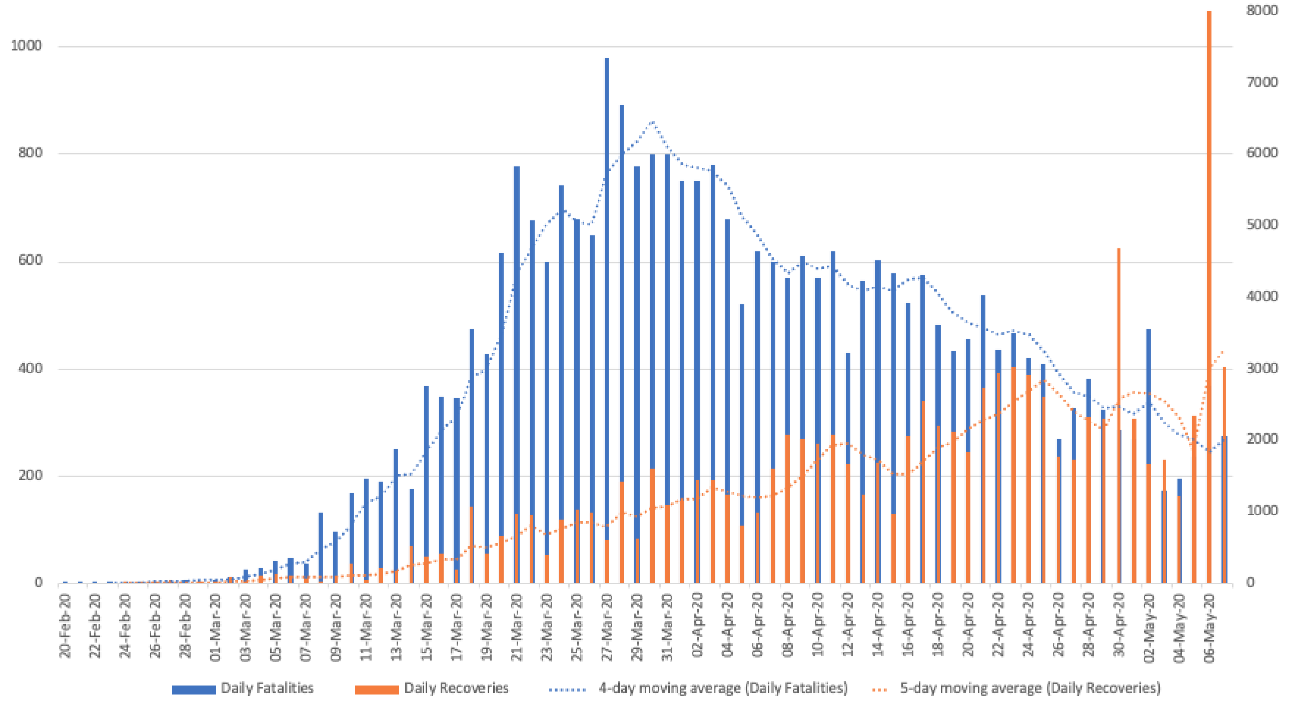


Figure 2. Graphical plot of COVID-19 cumulative total fatalities versus recoveries in Italy. COVID-19: coronavirus disease.

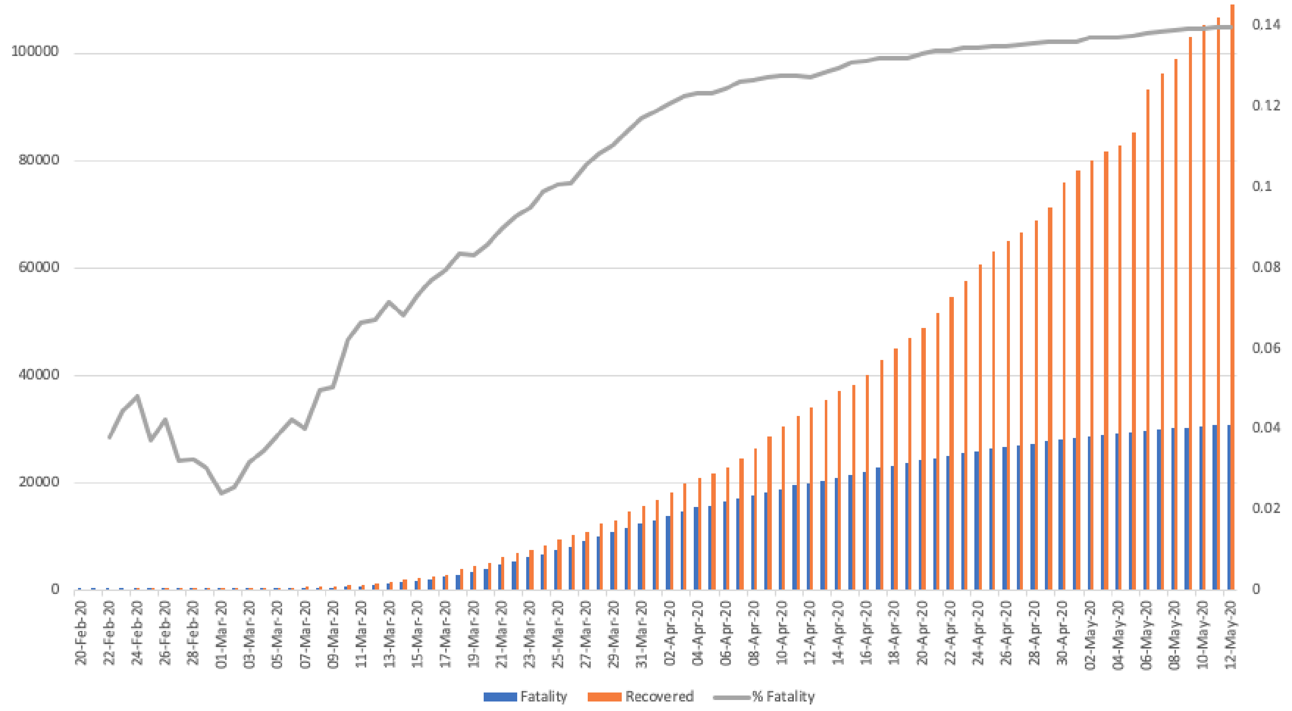


Table 1 shows the COVID-19 case fatality rates and mortality rates (the ratio of deaths of patients who tested positive for COVID-19 to the total number of citizens) on April 14, 2020 among the industrialized countries in the G10: Belgium, Canada, France, Germany, Japan, Italy, the Netherlands, the United Kingdom, the United States, and Sweden [5]. Comparing the Italian data to those of the other G10 countries, the COVID-19 mortality in Italy is approximately 358 per 1 million inhabitants, which is not the highest value in the set. Belgium, for instance, has a mortality rate of 419 per 1 million inhabitants, and the rate in Spain is 409 per 1 million inhabitants. Following the model of the statistical exercise above, if we excluded the Lombardy region mortality rate of 1000 per 1 million inhabitants, the Italian mortality rate would decrease to 220 per 1 million inhabitants.

Table 1. Comparison of the case fatality rate and mortality rate in Italy and the Lombardy region with those of the other G10 countries.

Country	Case fatality rate, %	Mortality rate, deaths per 1 million inhabitants
Belgium	12.80	419
Canada	4.00	32
France	11.20	263
Germany	2.70	46
Italy	12.80	358
Japan	2.12	2
Lombardy region	18.08	1,000
Netherlands	11.40	193
United Kingdom	12.70	190
United States	5.50	86
Sweden	10.50	132
Spain	10.44	409

COVID-19 Test Coverage in Italy

One of the most debated issues in Italy is the number of COVID-19 tests performed [1]; the scientific community is divided by suggestions of a hypothetical high number of undiagnosed cases that would lower the national case fatality rate. This aspect can be easily assessed from the officially declared world data [5]. As reported in Table 2, at the time of writing (April 15, 2020), approximately 1,100,000 COVID-19 tests had been performed in Italy. In the United States, where over 600,000 COVID-19 cases had been confirmed to date, 3,300,000 tests were performed. In Spain (the second most affected country in the world, with over 170,000 total cases), only 600,000 tests were performed; 330,000 tests were performed in France, 1,300,000 in Germany, and 380,000 in the United Kingdom. Therefore, in absolute terms, the nation of Italy has performed the third largest number of tests in the world.

After weighting the number of tests by the population of each nation, Italy has a ratio of 17,800 tests per 1 million inhabitants, ranking first among the G10 countries in terms of tests per 1 million inhabitants (column 3 of Table 2) [5]. This gap is even more evident for the Lombardy region, which has a ratio of 25,000 tests per 1 million inhabitants.

Conversely, if we consider the total of patients who tested positive and the number of swabs performed in the equation, we find that the lower the percentage ratio, the more effectively the nation has tested its citizens.

It is important to observe the temporal misalignment between all the countries because the COVID-19 outbreak started in different countries on different dates. Hence, these values should be normalized by the cumulative number of days since the first outbreak in each country. However, it should be noted that the Lombardy region suddenly discovered the first outbreak with a delay of at least 18 days (ie, 2 weeks of average SARS-CoV-2 incubation time plus 4 or 5 days to develop severe symptoms). In contrast, other European Union countries were able to prepare in advance and may have benefited from a significant advantage in managing the diffusion of COVID-19 by observing the behavior of the COVID-19 spread in Italy, particularly in the Lombardy region.

Table 2. COVID-19 testing data for Italy and the Lombardy region and for the other G10 countries.

Country	Total tests	Tests per 1 million inhabitants	Positive cases/tests, %
Belgium	134,000	11,000	26.0
Canada	450,000	12,000	6.7
France	330,000	5,000	44.3
Germany	1,300,000	16,900	8.0
Japan	100,000	800	9.0
Italy	1,100,000[a]	17,800[b]	14.8[c]
Netherlands	140,000	8,500	20.0
United Kingdom	380,000	6,000	24.6
United States	3,300,000	10,000	19.8
Sweden	70,000	7,000	16.6
Lombardy region	230,000	23,000	27.9
Spain	600,000	17,500	19.7

[a] Third place.

[b] First place.

[c] Fourth place.

Additional Considerations in the Lombardy Region

As already discussed, the Lombardy region has a very significant impact on the national data in Italy. Hence, it may be worthwhile to explore additional considerations to better understand the impact of the Lombardy region on the global data. The Lombardy region has 10,000,000 inhabitants in a territory of 23,800 square kilometers and has a population density of 425 inhabitants/km^2. Its capital city, Milan, is one of the most dynamic alpha cities in the world, with a population of 1,350,000 people and a population density that reaches 7600 units/km^2. The Lombardy region also contains the largest logistics hubs in Italy (close to the first outbreak in Codogno [5]) and some of the largest in Europe. It is also one of the regions in Europe with the greatest flow of people and goods by air, rail, or road [10].

In this context, the analysis can be further extended by comparing the Lombardy region data with the most important regions/areas of the other G10 countries (see Table 3). If we focus on the total cases reported, the Lombardy region is the region with the second highest absolute value of COVID-19–positive cases (Lombardy region=59,000; New York City=123,000). If we normalize this value by the total number of inhabitants, New York, Antwerp, and Madrid have the highest numbers of positive cases per 1 million inhabitants. Regarding case fatality rates, the Lombardy region is the region with the highest number of fatalities (18.08%), followed by Île-de-France (17.99%) and Antwerp (14.20%). When we normalize the regional case fatality rate with the national datum (ie, the ratio of the regional case fatality rate to the national case fatality rate), New York City appears to have the highest multiplication factor (9.50%/5.50%) of ×1.73.

Table 3. COVID-19 statistics for the Lombardy region and for the most affected regions in the other G10 countries.

Country	Region	Total cases	Cases per 1 million inhabitants	Case fatality rate, %	Ratio of regional case fatality rate to national case fatality rate
France	Île-de-France	23,500	2000	17.00	1.52
Germany	Bayern	36,000	2800	3.20	1.18
Italy	Lombardy region	59,000	5900	18.08	1.41
Belgium	Antwerp	4500	9500	14.20	1.25
United Kingdom	London	19,500	2200	—[a]	—
Sweden	Stockholm County	7100	7500	14.10	1.34
Spain	Community of Madrid	48,000	8000	14.00	1.34
United States	New York City	123,000	13,700	9.50	1.73

[a] Not available.

Discussion

The case fatality rate and mortality rate of COVID-19 in Italy are in line with the official world scenario. It is also evident that the numbers of COVID-19 tests performed in Italy and in the Lombardy region are comparable with those in the rest of Europe and the rest of the world.

Acknowledgments

The work presented here has been partly supported by the FP7 Collaborative Project S-CASE (Grant Agreement No. 610717), funded by the European Commission.

Conflicts of Interest

None declared.

References

1. Italian Department of Civil Protection. Coronavirus emergency. Webpage in Italian URL: http://www.protezionecivile.gov.it/ [accessed 2020-03-22]
2. Paterlini M. On the front lines of coronavirus: the Italian response to covid-19. BMJ 2020 Mar 16;368:m1065. [doi: 10.1136/bmj.m1065] [Medline: 32179517]
3. Anderson M, Mckee M, Mossialos E. Covid-19 exposes weaknesses in European response to outbreaks. BMJ 2020 Mar 18;368:m1075. [doi: 10.1136/bmj.m1075] [Medline: 32188590]
4. Rudan I. A cascade of causes that led to the COVID-19 tragedy in Italy and in other European Union countries. J Glob Health 2020 Jun;10(1):010335 [FREE Full text] [doi: 10.7189/jogh-10-010335] [Medline: 32257150]
5. Worldometer. Official World-Wide COVID-19 Data URL: https://www.worldometers.info/coronavirus/ [accessed 2020-03-22]
6. Government of Italy. 2020 Mar 22. Decree of the President of the Italian Council of Ministers of 22 March 2020 URL: https://www.assolombarda.it/servizi/internazionalizzazione/documenti/dpcm-22-marzo-2020-in-inglese [accessed 2020-06-03]
7. Tosi D. ResearchGate. 2020 Mar 08. Studio di un Modello di Predizione per i Contagi da COVID-19 in Italia e Regione Lombardia URL: https://www.researchgate.net/publication/340090910_Studio_di_un_Modello_di_Predizione_per_i_Contagi_da_COVID-19_in_Italia_e_Regione_Lombardia [accessed 2020-06-03]
8. Robert Koch Institute. 2020 Apr 14. Coronavirus Disease 2019 (COVID-19) Daily Situation Report of the Robert Koch Institute URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-04-14-en.pdf?__blob=publicationFile&fbclid=IwAR0b66TJBHnuuIYTzaDicnN5i8g2J5kEw4yYkZ0nw5GhDgYM5Vll6o5Ka38 [accessed 2020-06-03]
9. Global Health 5050. COVID-19 sex-disaggregated data tracker URL: https://globalhealth5050.org/covid19/ [accessed 2020-06-03]
10. European Commission. Coronavirus Response URL: https://ec.europa.eu/info/index_en [accessed 2020-06-03]

Abbreviations

COVID-19: coronavirus disease
csv: comma-separated value
G10: Group of Ten
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 03.05.20; peer-reviewed by G Aiello, K Goniewicz; comments to author 27.05.20; revised version received 30.05.20; accepted 03.06.20; published 17.06.20.

Please cite as:
Tosi D, Verde A, Verde M
Clarification of Misleading Perceptions of COVID-19 Fatality and Testing Rates in Italy: Data Analysis
J Med Internet Res 2020;22(6):e19825
URL: http://www.jmir.org/2020/6/e19825/
doi:10.2196/19825
PMID:32490842

Original Paper

Human Factor Considerations in Using Personal Protective Equipment in the COVID-19 Pandemic Context: Binational Survey Study

Avi Parush[1*], PhD; Oren Wacht[2*], EMT-P, PhD; Ricardo Gomes[3], MD, OF-4 GNR; Amit Frenkel[4], MD, MHA

[1]Israel Institute of Technology, Haifa, Israel

[2]Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel

[3]Department of Operational Medicine, Clinical Center of Guarda Nacional Republicana, Lisboa, Portugal

[4]Intensive Care Unit, Soroka University Medical Center, Beer Sheva, Israel

[*]these authors contributed equally

Corresponding Author:
Avi Parush, PhD
Israel Institute of Technology
Technion City
Haifa, 3200003
Israel
Phone: 972 0502301857
Email: aparush@technion.ac.il

Abstract

Background: Full level 1 personal protective equipment (PPE) is used in various domains and contexts. Prior research has shown influences of such equipment on performance, comfort, and contamination levels. The coronavirus disease (COVID-19) pandemic forced a pervasive requirement of PPE, with little preparation, rushed deployment, inadequate time for training, and massive use by personnel who are inexperienced or not qualified in its effective use.

Objective: This study aims to examine the key human factors (physical and ergonomic, perceptual and cognitive) that influence the use of level 1 PPE when attending to patients with suspected or confirmed COVID-19.

Methods: The research approach consisted of a short survey disseminated to health care professionals in two countries, Israel and Portugal, with similar demographics and health care systems. The survey included 10 items with a 5-point Likert scale regarding the key human factors involved in level 1 PPE, as identified in prior research.

Results: A total of 722 respondents from Israel and 301 respondents from Portugal were included in the analysis. All the respondents reported using level 1 PPE with patients with COVID-19 in the range of several hours daily to several hours weekly. The Cronbach α was .73 for Israel and .75 for Portugal. Responses showed high levels of difficulty, with medians of 4 for items related to discomfort (n=539/688, 78% in Israel; n=328/377, 87% in Portugal), hearing (n=236/370, 64% in Portugal; n=321/642, 50% in Israel), seeing (n=697/763, 89% in Israel; n=317/376, 84% in Portugal), and doffing (n=290/374, 77% in Portugal; n=315/713, 44% in Israel). A factor analysis showed a set of strongly related variables consisting of hearing, understanding speech, and understanding the situation. This suggests that degradation in communication was strongly associated with degradation in situational awareness. A subsequent mediation analysis showed a direct effect of PPE discomfort on situational awareness (*P*<.001); this was also influenced (mediated) by difficulties in communicating, namely in hearing and understanding speech.

Conclusions: In 2020, the COVID-19 pandemic is paving the way for updating PPE design. The use of already deployed technology affords ample opportunities to improve, adapt, and overcome caveats. The findings here suggest that the use of level 1 PPE with patients with COVID-19 has perceptual and cognitive effects, in addition to physical and ergonomic influences. Efforts should be taken to mitigate the harmful effects of such influences, both regarding the performance of medical actions and the risk of contamination to health care workers. Such efforts involve the design of PPE; the introduction of technologies to enhance vision, hearing, and communicating during the use of PPE; and training staff in using the equipment and in effective communication and teamwork protocols.

(J Med Internet Res 2020;22(6):e19947) doi:10.2196/19947

KEYWORDS

COVID-19; personal protective equipment; PPE; human factors; cognitive functioning; multinational survey; pandemic; protection; infectious disease; infection; survey

Introduction

Starting in December 2019 and during the first months of the year 2020, the global outbreak of the coronavirus disease (COVID-19) has forced health care professionals of various disciplines in hospital and community settings to use full level 1 personal protective equipment (PPE) to avoid contamination from patients with suspected or confirmed COVID-19 [1]. Such PPE typically consists of a completely encapsulated suit and a self-contained breathing apparatus, such as the N95 face mask, which can provide full skin, eye, and respiratory protection.

The pervasive requirement to use PPE due to the COVID-19 pandemic emerged with little preparation and rushed deployment, inadequate time for training, and massive use by personnel who are inexperienced or not qualified in the effective use of PPE. Such unique and urgent circumstances call for an examination of the use of the full PPE. Moreover, the current widespread use of PPE is not limited to contact with patients with COVID-19 in dedicated and isolated units, but rather in a variety of contexts, and in the general community.

Previous research on using full PPE in various contexts (chemical, biological, radiological, nuclear and explosive incidents, firefighting, health care, and the military) has shown several human factor problems. Much of the research addressed procedural problems, including failures to effectively put on (don) or remove (dof) the PPE [2-5], and ergonomic problems such as poor fit and discomfort while having it on [6-9]. Problems of ineffective use or decreased adherence to using the PPE were found to be associated with insufficient training and lack of prior experience [10-13], and with the appropriateness of organizational culture [12,14-16].

The use of PPE can also influence perceptual and cognitive functioning, although less research has been conducted about such influences. PPE can degrade visual perception [17]; auditory perception [18,19]; gait and balance, which are related to vision and hearing [20,21]; and cognitive functioning [22,23], communication, and teamwork [8,17,24]. Finally, according to some research, the use of PPE can result in ineffective protection against contamination [25,26]. Moreover, even effective use of PPE can influence medical actions including lifesaving procedures [13,27-29].

Taken together the use of PPE, whether effective or ineffective, has been shown to influence the user's functioning and performance, as well as protection from contamination. Yet, the COVID-19 pandemic has introduced unique circumstances and challenges regarding the use of PPE, which prompt a critical need to re-examine the influences. This study explored these influences by means of a binational survey. In particular, we set out to identify relations of physical and ergonomic factors with perceptual and cognitive factors in the use of PPE during the COVID-19 pandemic. For this study, we considered two small developed countries that used equivalent PPE levels to deal with the COVID-19 pandemic: Israel and Portugal. The countries are similar in population, median age, and life expectancy, and have well-ranked national health care systems (NHS), 25th and 31st, respectively, in a 2018 report by the UK Health Foundation [30].

Methods

The Two Countries

Israel has an area of 22,145 km^2, a population of 9,190,000, and a current density of 400 inhabitants per km^2. The median age is 30.5 years and the life expectancy is 83.5 years. The Israeli Ministry of Health is responsible for managing the NHS by means of a public health system. The national emergency medical service is called Magen David Adom. Activated by the emergency number 101 under the coordination of the dispatch centers, specialized resources operate after triage of prehospital levels of care: basic or advanced life-support motorcycles (1 emergency medical technician [EMT] or 1 paramedic), basic life support ambulances (1 or 2 EMTs), intensive care ambulances or helicopters (1 or 2 paramedics and 1 EMT in an ambulance) [31].

Portugal has an area of 92,212 km^2, a population of 10,202,166, and a current density of 111 inhabitants per km^2. The median age is 46.2 years and the life expectancy is 81.5 years. The Portuguese Ministry of Health is responsible for managing the NHS, tendential and free for all residents. The main emergency medical service is managed by the National Medical Emergency Institute. Activated by the emergency number 112 under the coordination of the dispatch centers, specialized resources operate after triage of prehospital level of care: basic life support motorcycles (1 EMT), basic life support ambulances (2 EMTs), intermediate life support ambulances (1 EMT and 1 nurse), advanced critical care fast cars (1 physician and 1 nurse), and helicopter ambulances (1 physician and 1 nurse plus 2 pilots) [32].

Similar level 1 PPE is used in the two countries as is shown in Figure 1.

Table 1 presents the COVID-19 status in each country on May 1, 2020.

Figure 1. Typical level 1 personal protective equipment used in Portugal (left) and Israel (right).

Table 1. Coronavirus disease summary statistics for Israel and Portugal, as of May 1, 2020.

Statistic	Israel	Portugal
Total confirmed cases, n	16,101	25,351
New cases[a], n	58	306
Total deaths, n	225	1007
New deaths[a], n	1	18
Total recovered, n	9156	1647
Active cases, n	7023	22,697
Intensive care unit cases, n	103	154
Total tests, n	390,022	395,771

[a] During the last 24 hours, as of May 1, 2020.

Survey Design

The survey consisted of ten items relating to factors that have been found to have influence during the use of full PPE. The survey included general information about the objective of the study and a photo of a person wearing level 1 PPE.

Each item of the survey stated either the difficulty or ease of an influencing factor. Half the statements were worded as a difficulty and half as an easing influence so as to mitigate response bias. The statements were as follows in this order: (1) it is easy to put on (don) the full PPE; (2) wearing the full PPE is very uncomfortable; (3) it is hard to see everything around me while wearing the full PPE; (4) it is easy to hear sounds and speech around me while wearing the full PPE; (5) it is easy to understand what is said to me while wearing the full PPE; (6) it is easy to understand my surroundings while wearing the full PPE; (7) it is hard to think clearly while wearing the full PPE; (8) it is easy to make decisions while wearing the full PPE; (9) it is hard to remove (dof) the full PPE; and (10) it is very important to wear the full PPE.

Responses were according to a Likert scale ranging from 1, completely disagree, to 5, completely agree. The questionnaire included three demographic questions regarding gender, profession, and the frequency of using the full PPE. All statements and demographic questions were presented in Hebrew to participants in Israel and in Portuguese to participants in Portugal.

The survey was administered through the Qualtrics (Qualtrics International Inc) online platform. It was available for participants from April 12 in Israel and from April 16 in Portugal, to May 1, 2020. This platform ensured that there were no multiple entries from the same individual.

The study was approved by the Institutional Review Board of the Israel Institute of Technology, Haifa, Israel. Participants received a general introduction to the study and then presented with an informed consent form. In that form, participants were informed that the study does not pose any risk, that their participation is completely voluntary, and that they can quit at any time. In addition, they were informed that no personal data or identifying details were required nor collected, and that the data was secure and their participation remained anonymous. Participants had to select the "I Agree" option before entering the survey itself.

Recruitment

We used a combination of criterion-based purposive and convenience sampling. The criterion for the purposive sampling was prehospital and hospital health care professionals involved in the care of patients with COVID-19 and using level 1 PPE. Using professional email lists and closed professional WhatsApp and Facebook groups, we recruited 1013 participants in Israel. In a similar fashion, we recruited 519 participants in Portugal.

Statistical Analysis

Data Exclusion

Of the original 1013 persons who entered the survey in Israel, 189 did not proceed with responding and were thus excluded from the analyses. Of the original 519 who entered the survey in Portugal, 104 did not proceed with responding and were thus excluded from the analyses. We assume that many who consented to participate and entered the survey may have realized that the questions focused only on the use of level 1 PPE and elected not to continue. In addition, the goal of the study was to examine the influences of using the PPE based on extended and intensive use typical of the COVID-19 situation. Consequently, participants who reported rarely or never using the level 1 PPE may have responded to the questions based on their training or other sources but not necessarily based on personal experience and were thus excluded from the final survey analysis. In the Israeli sample, this amounted to 102 (12% of the participating respondents), and in the Portugal sample, this amounted to 114 (27% of the participating respondents).

Scoring the Responses

To achieve uniform direction of the responses on the 5-category Likert scale, the responses on items worded as a statement of ease were reversed. The objective was that all high responses would indicate greater difficulty and low responses would indicate greater ease. Specifically, the responses to item 1, 4, 5, 6, and 8 were reversed for the subsequent analyses.

Results

Respondent Statistics

A total of 722 respondents from Israel (of a total of 824 who participated) were included in the analysis. This sample consisted of 346 (48%) who reported using the level 1 PPE for at least a few hours daily and 376 (52%) who reported its use for at least a few hours weekly. This sample included 524 (72%) males and 198 (28%) females. In addition, 66 (9%) were physicians, 46 (6%) were from nursing professions, 299 (41%) were paramedics, 269 (37%) were medics, and 42 (7%) were of other occupations.

A total of 301 respondents from Portugal (of a total of 415 who participated) were included in the analysis. This sample consisted of 150 (49%) who reported using the level 1 PPE for at least a few hours daily and 151 (51%) who reported its use for at least a few hours weekly. This sample included 168 (55%) males and 133 (45%) females. In addition, 69 (24%) were physicians, 130 (43%) were from nursing professions (which in Portugal includes prehospital emergency work, parallel to the paramedics in Israel), 64 (21%) were medics, and 37 (12%) were of other occupations.

Reliability

To assess reliability of the questionnaire items, a Cronbach α was computed for each of the samples. Cronbach α for the sample from Israel was .73 and, for the sample from Portugal, .75. This reliability index indicates an acceptable internal consistency of the responses.

Overall Responses on the Questionnaire Items

Medians and their corresponding ranges were computed for each of the questionnaire items, stratified by health care profession, for each of the two countries. In addition, the percentage of respondents indicating an agreement and a strong agreement with a given questionnaire statement were computed. These percentages represented the proportion of respondents that expressed more difficulty with a given PPE factor. These statistics are presented in Table 2.

Table 2. Medians for each of the survey items, together with the proportions of respondents who agreed and strongly agreed with the difficulty, stratified by profession, for each of the countries.

Survey Item/Profession	Israel		Portugal	
	Score, median (range)	Participants, n/N (%)	Score, median (range)	Participants, n/N (%)
Difficulty of donning				
Medicine	2 (1-5)	18/67 (27)	3 (1-5)	53/96 (55)
Nursing	2 (1-5)	16/26 (38)	2 (1-5)	60/140 (43)
Paramedics	2 (1-5)	65/268 (24)	—[a]	—
Medics	2 (1-5)	76/264 (29)	2 (1-5)	27/74 (36)
Other	2 (1-5)	13/43 (30)	2.5 (1-5)	22/52 (42)
Discomfort				
Medicine	4 (1-5)	42/61 (69)	4 (1-5)	96/102 (94)
Nursing	4 (1-5)	32/45 (71)	4 (1-5)	133/150 (89)
Paramedics	4 (1-5)	224/264 (85)	—	—
Medics	4 (1-5)	208/280 (74)	4 (1-5)	39/51 (82)
Other	4 (1-5)	33/38 (87)	4 (1-5)	39/51 (76)
Difficulty in seeing				
Medicine	4 (2-5)	55/70 (78)	4 (1-5)	82/94 (87)
Nursing	4 (2-5)	38/44 (86)	4 (1-5)	139/153 (91)
Paramedics	4 (2-5)	280/300 (93)	—	—
Medics	4 (1-5)	266/303 (88)	4 (1-5)	55/73 (75)
Other	4 (1-5)	40/46 (87)	3.5 (1-5)	40/50 (77)
Difficulty in hearing				
Medicine	3 (1-5)	31/65 (48)	4 (1-5)	68/97 (70)
Nursing	4 (1-5)	27/39 (69)	4 (1-5)	105/146 (72)
Paramedics	3 (1-5)	132/262 (50)	—	—
Medics	3 (1-5)	100/259 (38)	3 (1-5)	37/75 (49)
Other	3 (1-5)	24/36 (67)	3 (1-5)	25/49 (51)
Difficulty in understanding speech				
Medicine	3 (1-5)	32/61 (52)	4 (1-5)	70/96 (73)
Nursing	4 (1-5)	28/37 (76)	4 (1-5)	109/144 (76)
Paramedics	3 (1-5)	135/251 (54)	—	—
Medics	3 (1-5)	106/258 (41)	3 (1-5)	34/75 (45)
Other	3 (1-5)	20/34 (59)	3 (1-5)	29/53 (55)
Difficulty in understanding the surroundings				
Medicine	3 (1-5)	25/52 (48)	4 (1-5)	73/96 (76)
Nursing	3 (1-5)	24/39 (61)	4 (1-5)	107/147 (73)
Paramedics	3 (1-5)	147/236 (62)	—	—
Medics	3 (1-5)	108/238 (45)	4 (1-5)	45/72 (62)
Other	3 (1-5)	21/34 (62)	3 (1-5)	26/48 (54)
Difficulty in thinking clearly				
Medicine	2 (1-5)	22/60 (37)	4 (1-5)	50/84 (59)
Nursing	3 (1-5)	24/41 (58)	4 (1-5)	84/123 (68)
Paramedics	3 (1-5)	121/245 (49)	—	—

Survey Item/Profession	Israel		Portugal	
	Score, median (range)	Participants, n/N (%)	Score, median (range)	Participants, n/N (%)
Medics	2 (1-5)	92/260 (35)	3 (1-5)	27/64 (42)
Other	3 (1-5)	16/38 (42)	2 (1-5)	19/48 (40)
Difficulty in making decisions				
Medicine	2.5 (1-5)	14/51 (27)	3 (1-5)	50/79 (63)
Nursing	3 (1-5)	15/33 (45)	3 (1-5)	64/119 (54)
Paramedics	3 (1-5)	107/208 (51)	—	—
Medics	3 (1-5)	69/226 (30)	2 (1-5)	25/67 (37)
Other	3 (1-5)	13/34 (38)	2.5 (1-5)	18/48 (37)
Difficulty of doffing				
Medicine	3 (1-5)	38/70 (54)	4 (1-5)	80/93 (86)
Nursing	4 (2-5)	25/44 (57)	4 (1-5)	121/149 (81)
Paramedics	3 (1-5)	124/277 (45)	—	—
Medics	2 (1-5)	110/280 (40)	4 (1-5)	54/76 (71)
Other	2.5 (1-5)	18/42 (43)	4 (1-5)	34/52 (65)
The importance of wearing the personal protective equipment				
Medicine	5 (2-5)	74/75 (99)	5 (4-5)	104/107 (97)
Nursing	5 (2-5)	47/48 (98)	5 (4-5)	159/160 (99)
Paramedics	5 (1-5)	294/299 (98)	—	—
Medics	5 (1-5)	314/319 (98)	5 (4-5)	84/84 (100)
Other	5 (3-5)	46/46 (100)	5 (3-5)	60/60 (100)

[a] The emergency medical services in Portugal do not include paramedics; nurses working in prehospital settings fulfill parallel functions.

Responses to items relating to the physical and ergonomic aspects showed high agreement across professions and in both countries that the use of the PPE is highly uncomfortable: 78% (n=539/688) in Israel and 87% (n=328/377) in Portugal. Only 27% (n=188/684) of the respondents from Israel and 45% (n=163/365) from Portugal, of all the professions, indicated that donning the PPE was difficult. Agreement was high across professions and the two countries regarding difficulty in seeing what is going on around while using the PPE: 89% (n=697/763) in Israel and 84% (n=317/376) in Portugal. Finally, almost all the respondents of all the professions, 99% in both countries, expressed the importance of using the PPE.

Significantly higher proportions of respondents from Portugal than Israel reported difficulties in doffing the PPE: 77% (n=290/374) vs 44% (n=315/713; $\chi^2_1=110$, $P<.001$); in hearing: 64% (n=236/370) vs 50% (n=321/641; $\chi^2_1=25.2$, $P<.001$); in understanding speech: 65% (n=243/372) vs 47% (n=314/661; $\chi^2_1=22.1$, $P<.001$); in understanding the situation: 69% (n=253/367) vs 54% (n=325/599; $\chi^2_1=20.4$, $P<.001$); in being able to think clearly: 57% (n=183/323) vs 43% (n=275/644; $\chi^2_1=16.8$, $P<.001$); and in being able to make decisions: 50% (n=158/317) vs 39% (n=218/552; $\chi^2_1=8.8$, $P<.001$).

Discovery of Common Factors

When variables of a certain set correlate with each other, a certain factor may exist that correlates with the set of variables. To discover such possible underlying factors in the data, an exploratory factor analysis (EFA) was performed to reduce the data set into a smaller number of variables. Factor analysis identifies factors that may explain correlations among variables. Accordingly, each variable has a relative weight, or loading, within a given factor. Statistical indices of the analysis indicate that our data were suitable for factor analysis and suitable for the specific factor analysis reduction technique. The detailed statistics of the factor analysis for each country are included in Multimedia Appendix 1 for Israel and Multimedia Appendix 2 for Portugal.

The factor analysis uncovered two factors, which are presented for both countries in Table 3. For both countries, the first factor showed high loadings for three questionnaire items: difficulty in hearing, difficulty in understanding speech, and difficulty in understanding the situation. This factor suggests that when the full PPE is on, increased difficulties in hearing and understanding speech imply difficulties in spoken communication, which were related to difficulties in understanding the surroundings, that is, situational awareness.

Table 3. Relative weights (loading coefficients) of all the questionnaire items according to the two factors identified for each country[a].

Questionnaire items	Israel		Portugal	
	Factor 1	Factor 2	Factor 1	Factor 2
Donning	0.05	*0.65* [b]	0.29	0.53
Discomfort	0.33	0.56	0.19	*0.70*
Seeing	0.40	0.43	0.07	*0.71*
Hearing	*0.87*	0.02	*0.84*	0.07
Understanding speech	*0.89*	0.04	*0.87*	0.16
Understanding the surroundings	*0.64*	0.36	*0.80*	0.13
Thinking clearly	0.11	*0.63*	0.10	0.55
Making decisions	0.42	0.45	0.44	0.18
Doffing	0.02	*0.61*	0.06	*0.67*

[a] The analysis did not include the last item of the questionnaire, regarding the significance of having the personal protective equipment, since there was little variation in the responses to that question.

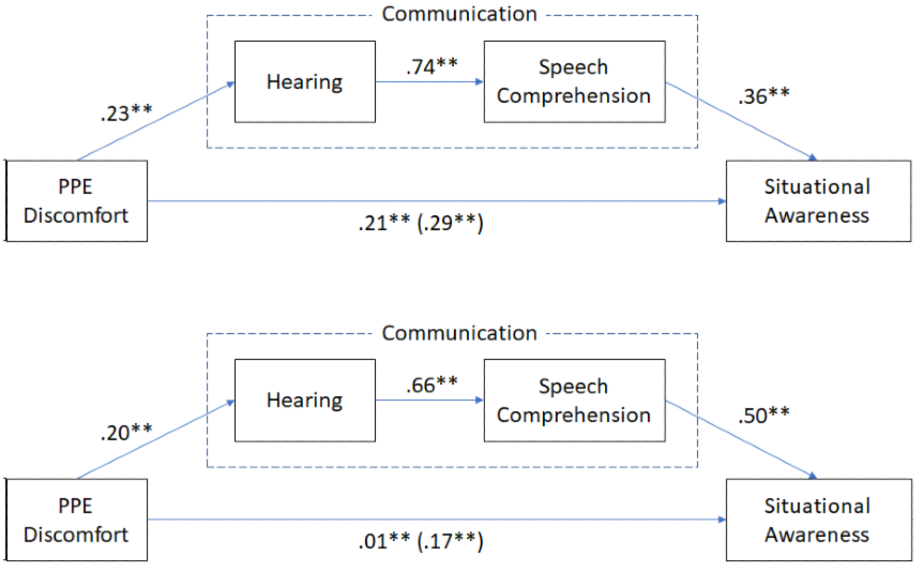
[b] Relative weight values higher than 0.6 are in italics.

Less similarity was observed between the two countries in the second factor. The Israeli data show loadings slightly higher than 0.6 for the items of difficulty of donning the PPE, difficulty in thinking clearly, and difficulty of doffing, with discomfort being close to a loading of 0.6. The Portuguese sample showed loadings higher than 0.6 for the items of discomfort, difficulty in seeing, and difficulty of doffing. Taken together, this factor implies a construct relating to the physical aspects of having the PPE, primarily related to doffing, discomfort, and donning. Interestingly, the second factor in the Portuguese sample includes the item of difficulty in seeing.

PPE Discomfort, Communication, and Situational Awareness: Relations Between the Constructs

We further explored the possibility that wearing the PPE influences communication, as manifested through difficulties in hearing and understanding speech, which in turn influences situational awareness. Specifically, we examined the extent that discomfort influenced communication (ie, hearing and understanding speech), which in turn influenced situational awareness. We employed a mediation analysis based on [33], using the Process Macro in SPSS v21 (IBM Corp). Specifically, we used model 6, which assumes two mediators between an independent variable and the outcome. In this study, the extent of difficulties in hearing while having the PPE on was the independent variable, and the outcome was the extent that one understands the surroundings. We explored the extent of hearing and understanding what is spoken while wearing the PPE, as possible mediating variables in the influence of PPE discomfort on situational awareness. This mediation analysis was performed on the data of each country. Both mediation analyses are presented in Figure 2.

Figure 2. Mediation models of the responses in Israel (top model) and Portugal (bottom model). Each link is associated with its corresponding standardized coefficient, together with the significance indication (** indicates a significance level of $P<.01$). The direct link between PPE discomfort and situational awareness shows the total effect, including the presence of the mediating variables; the direct effect without the mediating factors is in parenthesis. PPE: personal protective equipment.



As shown in Figure 2, discomfort had a significant direct effect on situational awareness, as evident by the significant coefficients in parenthesis, for both countries. This effect decreased significantly when mediation by hearing and understanding speech was considered. This implies a significant indirect effect of discomfort on situational awareness, as mediated by hearing and understanding speech. Increased discomfort with the PPE was related to increased difficulties in hearing and speech comprehension, which were related to increased difficulties in understanding the surroundings. According to this analysis, the mediating variables of hearing and understanding speech could account for 31% of the total effect of discomfort on situational awareness in Israel and 94% in Portugal. Although the effect size in the Portuguese data was much stronger than in the Israeli data, the same relations were demonstrated significantly in both countries.

Discussion

Principal Results

During the period this study was conducted, level 1 PPE was used extensively to mitigate contamination of health care professions treating patients with suspected or confirmed COVID-19. A survey of health care workers in two countries, Israel and Portugal, showed that the use of level 1 PPE is uncomfortable and entails difficulties in removing (doffing) it. Moreover, the discomfort appeared to be associated with difficulties in perceptual and cognitive functioning. Specifically, the equipment was found to be associated with difficulties in visual and auditory perception, and difficulties in understanding speech, situational awareness, and thought and decision making.

The findings of the survey uncovered strong relations between difficulties in auditory perception, the ability to understand speech, and situational awareness. Auditory perception and the ability to understand what is being said are fundamental and critical to speech communication. Communication is critical to teamwork [34] and to situational awareness [35]. Additional analysis of the survey responses shows associations of increased discomfort with the PPE with greater difficulties in communication, namely hearing and understanding speech; these in turn were associated with degraded situational awareness.

The mediation of communication in the influence of PPE discomfort on situational awareness has both interesting theoretical and practical implications. Theoretically, this implies that perceptual and cognitive processes play a significant role in the influence of a physical factor, discomfort in this case, and on a cognitive factor, situational awareness in this case. However, beyond these theoretical implications, practical implications can and should be drawn from such findings.

Practical Implications

Effective design of PPE and adequate sequence and training of donning and doffing are aimed at two major objectives: (1) protect health care workers from contamination; and (2) minimize interference with performing medical and nursing tasks. This section discusses practical implications of the similarities and differences in the responses between the two countries in terms of PPE design, donning and doffing procedures, training, and culture. It also addresses the findings in terms of short- and long-term implementations.

The results showed overall agreement across professions and the two countries concerning the discomfort of the PPE. Accounting for these similar responses is the fact that the components of the PPE, and the steps and sequences of donning and doffing PPE are similar for Israel and Portugal, with both following the CDC and WHO recommendations [36]. The donning sequence is: (1) gown or coverall; (2) mask or respirator; (3) goggles; (4) face shield; and (5) gloves. The doffing sequence is the reverse. Improper use of PPE is known to spread infection among health care workers [4,5]. All these steps can be sources of errors. This is especially true in the context of long work shifts and harsh environmental settings, as the design of PPE materials does not yet account for user temperature and hydration homeostasis. Both reusable and disposable PPE should be clearly designated such as with color coding and designed for easy donning and doffing [37,38]. The simpler the system (with fewer parts), the more likely it will avoid human distractions that may lead to contamination. Regarding contamination protection, although PPE is often only worn for short periods of time, pathogenic viruses such as influenza, severe acute respiratory syndrome, and Ebola can survive for extended periods of time on surfaces and be sources of transmission via surface-to-hand and hand-to-face or -mucous membrane contact. Despite two layers of protective clothing and sometimes two pairs of gloves, hand hygiene remains an essential aspect of PPE.

Full-face covers can provide adequate mucous membrane protection, such as the nasal cavities, lips, and mouth [39]. Whether by using positive pressure systems or only barrier filters, we believe that full-face integrated systems may widen the field-of-view and offer advantages for situational awareness. Such system should be designed to minimize condensation. In addition, a single system is simpler to maintain and minimizes efforts in donning and doffing.

Much of the research on influences of using PPE has focused on influences of the equipment on the motoric aspects of user performance [27,29,40]. The findings here suggest that the use of level 1 PPE also has perceptual and cognitive influences. Relating to communication and situational awareness, some key technological directions and advancements can be considered, such as audio and communication-enhancing technologies. These should be hands-free technologies under the PPE, such as headsets and microphones or throat microphones based on Bluetooth, Wi-Fi, or radio ultra-high frequency. These can be interfaced with dedicated mobile phone apps and facilitate effective hearing and talking among health care workers.

Longer-term consideration of technologies that promote more effective use of level 1 PPE and that facilitate perceptual and cognitive functioning can include:

- Virtual reality technology: the use of virtual reality for training purposes, such as was previously proposed [41], can facilitate training in effective communication,

teamwork, and situational awareness during the use of the equipment.
- Augmented reality technology: this single key aspect may boost the ergonomics of a full-face system and act synergistically when connected with a hands-free voice-activated communication device.
- Internet of Things (IoT): capable 5G IoT devices coupled with mesh framework solutions and cheap microcomputers (or even smartphones) can contribute to a solution. Ultrafast transmission rates and processing power, allied with powerful cloud servers, will eliminate bandwidth problems and eventually clear the way for artificial intelligence “Wingmen.” This should help overcome the previously mentioned limitations in PPE while providing advanced advice immediately [42].

The responses from Portugal reflected more difficulties in hearing, understanding speech, and understanding the situation in addition to more difficulties in thinking clearly and making decisions compared with the responses from Israel. These differences could be accounted for by the differences in training and overall culture. Specifically, the prehospital professionals in Israel, paramedics and medics, undergo more extensive training in using PPE as an overall preparedness to various chemical, biological, radiological, nuclear, and explosive materials incidents. This is in addition to the training at the breakout of the COVID-19 crisis. In contrast, the health care professionals in Portugal received training only when the crisis started. Thus, beyond considering the redesign of level 1 PPE and implementation of newer technologies to facilitate visual and auditory perception, more periodic training and procedures related to cognitive functioning can be implemented immediately. These can include practicing and adopting visual scanning patterns, and communication protocols with coworkers to facilitate teamwork situational awareness.

Another factor that could account for some of the differences between the countries are cultural and overall political considerations. Israel thus far is in a constant state of preparedness to multi-casualty incidents, and consequently health care professionals are in a different mindset when encountering a crisis compared with other countries. Thus, when dealing with a global crisis such as the COVID-19 pandemic, cultural and political characteristics should be considered.

Limitations and Future Research

The research has some limitations that should be addressed in follow-up and future studies. First, the definition of a paramedic differed between the two countries. In Israel, paramedics are all prehospital emergency workers; in Portugal, the nursing profession includes prehospital and in-hospital emergency workers. This distinction was not captured in the data and analyses. Second, the sampling method was indeed purposive but also comprised convenience sampling through the professional networks in each of the countries. This sampling approach may not have included other health care workers who used level 1 PPE during the pandemic, such as community health care workers. Such data would have enriched the data, findings, and conclusions. Third, the survey instructions specified the details of level 1 PPE, including a photo to visualize it. Nevertheless, we acknowledge the availability of a variety of level 1 PPE in the world, of different brands and including different components. The exact PPE used by each respondent was not captured in the data and should be documented and analyzed in future studies.

Future studies should include simulation-based experiments in which various PPE designs and procedures can be compared in terms of their influence on potential contamination and on physical, perceptual, and cognitive functioning. In addition, analytic techniques such as failure mode and effects analysis along with the risk priority ratings should accompany empirical studies.

Conclusions

From 2014 to 2016, a global epidemic of the Ebola virus made proper use of PPE a paramount concern in health care settings [43]. Nevertheless, almost no innovations were implemented in terms of PPE ergonomics to address limitations in hearing, vision, or even comfort. The COVID-19 pandemic serves as a “field test” for PPE technology, as every health care system in the world must adapt its use on a daily basis. Our binational study found that health care workers from different countries and cultures, yet similar NHS and PPE level approaches, shared the same difficulties and felt their situational awareness undermined by outdated PPE designs.

In 2020, the COVID-19 pandemic is paving the way for updating PPE design. The use of already deployed technology provides abundant opportunities to improve, adapt, and overcome problems. This should be done in a practical and aesthetically pleasing way, as well as appropriately to endure future epidemics of this century.

Acknowledgments

Each author has contributed substantially to the following: conception and design of the study or acquisition of data or analysis and interpretation of data; revising the article for important intellectual content; and final approval of the version submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Factor analysis statistics for Israel.

[PDF File (Adobe PDF File), 121 KB - jmir_v22i6e19947_app1.pdf]

Multimedia Appendix 2
Factor analysis statistics for Portugal.
[PDF File (Adobe PDF File), 120 KB - jmir_v22i6e19947_app2.pdf]

References

1. World Health Organization. 2020. Rational use of personal protective equipment (PPE) for coronavirus disease (COVID-19): interim guidance, 19 March 2020 URL: https://apps.who.int/iris/handle/10665/331498
2. Akbar-Khanzadeh F, Bisesi MS, Rivas RD. Comfort of personal protective equipment. Appl Ergonomics 1995 Jun;26(3):195-198. [doi: 10.1016/0003-6870(95)00017-7]
3. Mumma J, Durso F, Casanova L, Erukunuakpor K, Kraft CS, Ray SM, et al. Common behaviors and faults when doffing personal protective equipment for patients with serious communicable diseases. Clin Infect Dis 2019 Sep 13;69(Suppl 3):S214-S220 [FREE Full text] [doi: 10.1093/cid/ciz614] [Medline: 31517977]
4. Hallihan G, Baers J, Wiley K, Davies J, Kaufman J, Conly J, et al. Human factors evaluation to identify systems factors to improve safety during donning and doffing personal protective equipment (PPE) in Ebola virus disease management scenario. Open Forum Infect Dis 2015;2:1398. [doi: 10.1093/ofid/ofv133.952]
5. Gurses AP, Dietz AS, Nowakowski E, Andonian J, Schiffhauer M, Billman C, CDC Prevention Epicenter Program. Human factors-based risk analysis to improve the safety of doffing enhanced personal protective equipment. Infect Control Hosp Epidemiol 2019 Feb;40(2):178-186. [doi: 10.1017/ice.2018.292] [Medline: 30520708]
6. Akbar-Khanzadeh F. Factors contributing to discomfort or dissatisfaction as a result of wearing personal protective equipment. J Human Ergol 1998;27:70-75. [doi: 10.11183/jhe1972.27.70]
7. de Almeida RACDS, Veiga MM, de Castro Moura Duarte FJ, Meirelles LA, Veiga LBE. Thermal comfort and personal protective equipment (PPE). Work 2012;41:4979-4982. [doi: 10.3233/wor-2012-0042-4979]
8. Grugle NL, Kleiner BM. Effects of chemical protective equipment on team process performance in small unit rescue operations. Appl Ergon 2007 Sep;38(5):591-600. [doi: 10.1016/j.apergo.2006.08.003] [Medline: 17049477]
9. Rebmann T, Carrico R, Wang J. Physiologic and other effects and compliance with long-term respirator use among medical intensive care unit nurses. Am J Infect Control 2013 Dec;41(12):1218-1223 [FREE Full text] [doi: 10.1016/j.ajic.2013.02.017] [Medline: 23768438]
10. John A, Tomas ME, Cadnum JL, Mana TS, Jencson A, Shaikh A, et al. Are health care personnel trained in correct use of personal protective equipment? Am J Infect Control 2016 Jul 01;44(7):840-842. [doi: 10.1016/j.ajic.2016.03.031] [Medline: 27181222]
11. Kwon JH, Burnham CD, Reske KA, Liang SY, Hink T, Wallace MA, et al. Assessment of healthcare worker protocol deviations and self-contamination during personal protective equipment donning and doffing. Infect Control Hosp Epidemiol 2017 Sep;38(9):1077-1083 [FREE Full text] [doi: 10.1017/ice.2017.121] [Medline: 28606192]
12. Nichol K, McGeer A, Bigelow P, O'Brien-Pallas L, Scott J, Holness DL. Behind the mask: determinants of nurse's adherence to facial protective equipment. Am J Infect Control 2013 Jan;41(1):8-13 [FREE Full text] [doi: 10.1016/j.ajic.2011.12.018] [Medline: 22475568]
13. Coates M, Jundi A, James M. Chemical protective clothing; a study into the ability of staff to perform lifesaving procedures. J Accid Emerg Med 2000 Mar;17(2):115-118 [FREE Full text] [doi: 10.1136/emj.17.2.115] [Medline: 10718233]
14. Cavazza N, Serpe A. Effects of safety climate on safety norm violations: exploring the mediating role of attitudinal ambivalence toward personal protective equipment. J Safety Res 2009;40(4):277-283. [doi: 10.1016/j.jsr.2009.06.002] [Medline: 19778651]
15. Olson R, Grosshuesch A, Schmidt S, Gray M, Wipfli B. Observational learning and workplace safety: the effects of viewing the collective behavior of multiple social models on the use of personal protective equipment. J Safety Res 2009 Oct;40(5):383-387. [doi: 10.1016/j.jsr.2009.07.004] [Medline: 19932320]
16. Neves HCC, Souza ACSE, Medeiros M, Munari DB, Ribeiro LCM, Tipple AFV. Safety of nursing staff and determinants of adherence to personal protective equipment. Rev Lat Am Enfermagem 2011;19(2):354-361 [FREE Full text] [doi: 10.1590/s0104-11692011000200018] [Medline: 21584383]
17. AlGhamri AA, Murray SL, Samaranayake VA. The effects of wearing respirators on human fine motor, visual, and cognitive performance. Ergonomics 2013;56(5):791-802. [doi: 10.1080/00140139.2013.767383] [Medline: 23514088]
18. Gaston JR, Mermagen T, Foots A, Dickerson K. Auditory localization performance in the azimuth for tactical communication and protection systems. J Acoust Soc Am 2015 Apr;137(4):2322-2322. [doi: 10.1121/1.4920476]
19. Scharine A, Letowski T, Sampson JB. Auditory situation awareness in urban operations. J Milit Strategic Stud 2009:1-24 [FREE Full text]
20. Wade L, Weimar W, Davis J. Effect of personal protective eyewear on postural stability. Ergonomics 2004 Dec;47(15):1614-1623. [doi: 10.1080/00140130410001724246] [Medline: 15545235]

21. Brown MN, Char RMML, Henry SO, Tanigawa J, Yasui S. The effect of firefighter personal protective equipment on static and dynamic balance. Ergonomics 2019 Sep;62(9):1193-1201. [doi: 10.1080/00140139.2019.1623422] [Medline: 31204598]
22. Wade C, Davis J, Marzilli TS, Weimar WH. Information processing capacity while wearing personal protective eyewear. Ergonomics 2006 Aug 15;49(10):955-967. [doi: 10.1080/00140130600557926] [Medline: 16803726]
23. Park J. The adverse impact of personal protective equipment on firefighters' cognitive functioning. Res J Costumec 2019 Feb 28;27(1):1-10. [doi: 10.29049/rjcc.2019.27.1.001]
24. Gupta D, Kumar S. WebmedCentral.com. 2018. Video phone communication with patients In isolation URL: http://www.webmedcentral.com/article_view/5461 [accessed 2020-04-16]
25. Salehi H, Pennathur PR, Da Silva JP, Herwaldt LA. Examining health care personal protective equipment use through a human factors engineering and product design lens. Am J Infect Control 2019 May;47(5):595-598. [doi: 10.1016/j.ajic.2018.10.008] [Medline: 30522839]
26. Kwon JH, Burnham CD, Reske KA, Liang SY, Hink T, Wallace MA, et al. Assessment of healthcare worker protocol deviations and self-contamination during personal protective equipment donning and doffing. Infect Control Hosp Epidemiol 2017 Sep;38(9):1077-1083 [FREE Full text] [doi: 10.1017/ice.2017.121] [Medline: 28606192]
27. Castle N, Bowen J, Spencer N. Does wearing CBRN-PPE adversely affect the ability for clinicians to accurately, safely, and speedily draw up drugs? Clin Toxicol (Phila) 2010 Jul;48(6):522-527. [doi: 10.3109/15563650.2010.491483] [Medline: 20550425]
28. Wiyor HD, Coburn JC, Siegel KL. Impact of clinician personal protective equipment on medical device use during public health emergency: a review. Disaster Med Public Health Prep 2020 Apr;14(2):273-283. [doi: 10.1017/dmp.2019.52] [Medline: 31397260]
29. Johnson AT. Respirator masks protect health but impact performance: a review. J Biol Eng 2016;10:4 [FREE Full text] [doi: 10.1186/s13036-016-0025-4] [Medline: 26865858]
30. Dayan M, Ward D, Gardner T, Kelly E. The Health Foundation. 2018 Jun. NHS at 70: How Good is the NHS? URL: https://www.health.org.uk/sites/default/files/NHS-70-How-good-is-the-NHS.pdf
31. Central Bureau of Statistics. URL: https://www.cbs.gov.il/EN/pages/default.aspx
32. Central Intelligence Agency. The World Factbook URL: https://www.cia.gov/library/publications/resources/the-world-factbook/
33. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: Guilford Publications; 2017.
34. Gluyas H. Effective communication and teamwork promotes patient safety. Nurs Stand 2015 Aug 05;29(49):50-57. [doi: 10.7748/ns.29.49.50.e10042] [Medline: 26243123]
35. Parush A, Kramer C, Foster-Hunt T, Momtahan K, Hunter A, Sohmer B. Communication and team situation awareness in the OR: implications for augmentative information display. J Biomed Inform 2011 Jun;44(3):477-485. [doi: 10.1016/j.jbi.2010.04.002] [Medline: 20381642]
36. Centers for Disease Control and Prevention. Sequence for putting on personal protective equipment (PPE) URL: http://www.cdc.gov/hai/pdfs/ppe/ppe-sequence.pdf [accessed 2020-06-02]
37. Reddy S, Valderrama A, Kuhar D. Improving the use of personal protective equipment: applying lessons learned. Clin Infect Dis 2019 Sep 13;69(Suppl 3):S165-S170. [doi: 10.1093/cid/ciz619] [Medline: 31517978]
38. Drews F, Mulvey D, Stratford K, Samore M, Mayer J. Evaluation of a redesigned personal protective equipment gown. Clin Infect Dis 2019 Sep 13;69(Suppl 3):S199-S205 [FREE Full text] [doi: 10.1093/cid/ciz520] [Medline: 31517973]
39. Yan Y, Chen H, Chen L, Cheng B, Diao P, Dong L, et al. Consensus of Chinese experts on protection of skin and mucous membrane barrier for health professions fighting against coronavirus disease 2019. Chin J Dermatol 2020 Feb 20:e13310. [doi: 10.35541/cjd.20200112]
40. Loibner M, Hagauer S, Schwantzer G, Berghold A, Zatloukal K. Limiting factors for wearing personal protective equipment (PPE) in a health care environment evaluated in a randomised study. PLoS One 2019;14(1):e0210775 [FREE Full text] [doi: 10.1371/journal.pone.0210775] [Medline: 30668567]
41. Ragazzoni L, Ingrassia PL, Echeverri L, Maccapani F, Berryman L, Burkle FM, et al. Virtual reality simulation training for Ebola deployment. Disaster Med Public Health Prep 2015 Oct;9(5):543-546. [doi: 10.1017/dmp.2015.36] [Medline: 25782591]
42. Podgórski D, Majchrzycka K, Dąbrowska A, Gralewicz G, Okrasa M. Towards a conceptual framework of OSH risk management in smart working environments based on smart PPE, ambient intelligence and the Internet of Things technologies. Int J Occup Saf Ergon 2017 Mar;23(1):1-20. [doi: 10.1080/10803548.2016.1214431] [Medline: 27441587]
43. Centers for Disease Control and Prevention. 100 year of respiratory protection history URL: https://www.cdc.gov/niosh/npptl/Respiratory-Protection-history.html [accessed 2020-06-02]

Abbreviations

COVID-19: coronavirus disease
EFA: exploratory factor analysis

EMT: emergency medical technician
IoT: Internet of Things
NHS: national health care system
PPE: personal protective equipment

Edited by G Eysenbach; submitted 07.05.20; peer-reviewed by IC Lin, MB Uddin; comments to author 26.05.20; revised version received 02.06.20; accepted 02.06.20; published 17.06.20.

Please cite as:
Parush A, Wacht O, Gomes R, Frenkel A
Human Factor Considerations in Using Personal Protective Equipment in the COVID-19 Pandemic Context: Binational Survey Study
J Med Internet Res 2020;22(6):e19947
URL: http://www.jmir.org/2020/6/e19947/
doi:10.2196/19947
PMID:32511099

Original Paper

Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study

Jeremy F Huckins[1], PhD; Alex W daSilva[1], BSc; Weichen Wang[2], ME; Elin Hedlund[1], BA; Courtney Rogers[1], MEd; Subigya K Nepal[2], BSc; Jialing Wu[2], PhD; Mikio Obuchi[2], MA; Eilis I Murphy[1], BA; Meghan L Meyer[1], PhD; Dylan D Wagner[3], PhD; Paul E Holtzheimer[4,5], MD; Andrew T Campbell[2], PhD

[1]Department of Psychological and Brain Science, Dartmouth College, Hanover, NH, United States
[2]Department of Computer Science, Dartmouth College, Hanover, NH, United States
[3]Department of Psychology, Ohio State University, Columbus, OH, United States
[4]National Center for PTSD, White River Junction, VT, United States
[5]Department of Psychiatry, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States

Corresponding Author:
Jeremy F Huckins, PhD
Department of Psychological and Brain Science
Dartmouth College
HB6207
Hanover, NH, 03755
United States
Phone: 1 508 657 4825
Email: jeremy.f.huckins@dartmouth.edu

Related Article:

Companion article: https://www.jmir.org/2021/6/e28892

Abstract

Background: The vast majority of people worldwide have been impacted by coronavirus disease (COVID-19). In addition to the millions of individuals who have been infected with the disease, billions of individuals have been asked or required by local and national governments to change their behavioral patterns. Previous research on epidemics or traumatic events suggests that this can lead to profound behavioral and mental health changes; however, researchers are rarely able to track these changes with frequent, near-real-time sampling or compare their findings to previous years of data for the same individuals.

Objective: By combining mobile phone sensing and self-reported mental health data among college students who have been participating in a longitudinal study for the past 2 years, we sought to answer two overarching questions. First, have the behaviors and mental health of the participants changed in response to the COVID-19 pandemic compared to previous time periods? Second, are these behavior and mental health changes associated with the relative news coverage of COVID-19 in the US media?

Methods: Behaviors such as the number of locations visited, distance traveled, duration of phone usage, number of phone unlocks, sleep duration, and sedentary time were measured using the StudentLife smartphone sensing app. Depression and anxiety were assessed using weekly self-reported ecological momentary assessments of the Patient Health Questionnaire-4. The participants were 217 undergraduate students, with 178 (82.0%) students providing data during the Winter 2020 term. Differences in behaviors and self-reported mental health collected during the Winter 2020 term compared to previous terms in the same cohort were modeled using mixed linear models.

Results: During the first academic term impacted by COVID-19 (Winter 2020), individuals were more sedentary and reported increased anxiety and depression symptoms (*P*<.001) relative to previous academic terms and subsequent academic breaks. Interactions between the Winter 2020 term and the week of the academic term (linear and quadratic) were significant. In a mixed linear model, phone usage, number of locations visited, and week of the term were strongly associated with increased amount of COVID-19–related news. When mental health metrics (eg, depression and anxiety) were added to the previous measures (week of term, number of locations visited, and phone usage), both anxiety (*P*<.001) and depression (*P*=.03) were significantly associated with COVID-19–related news.

Conclusions: Compared with prior academic terms, individuals in the Winter 2020 term were more sedentary, anxious, and depressed. A wide variety of behaviors, including increased phone usage, decreased physical activity, and fewer locations visited, were associated with fluctuations in COVID-19 news reporting. While this large-scale shift in mental health and behavior is unsurprising, its characterization is particularly important to help guide the development of methods to reduce the impact of future catastrophic events on the mental health of the population.

(J Med Internet Res 2020;22(6):e20185) doi:10.2196/20185

KEYWORDS

COVID-19; depression; anxiety; mobile sensing; sedentary; phone usage; mental health; behavior; pandemic; app

Introduction

Coronavirus Disease

An outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease (COVID-19, also known as 2019-nCov), was first reported in Wuhan, China in December 2019, and SARS-CoV-2 was identified as a novel coronavirus in January 2020. On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a global pandemic; as of April 27, 2020, COVID-19 was responsible for over 200,000 deaths and 3,000,000 confirmed cases worldwide [1]. COVID-19 is not only a grave public health issue; it also carries severe political, economic, educational, and social ramifications. COVID-19 continues to impact millions of people worldwide every day. Understanding the behavioral and mental health implications for individuals during this unprecedented period of high stress and crisis is critical for informing current public policies and ensuring preparedness for future pandemics.

Mental Health and Behaviors in Pandemics and Disasters

Initial survey-based research on the psychological impact of the COVID-19 outbreak in China suggested that the mental health impact was moderate to severe for the majority of respondents in the general population, with increased anxiety, depression, and stress attributed to the outbreak by participants [2]. In initial research investigating the impact of the COVID-19 outbreak on college students in China, increased levels of anxiety and depression were observed as well as a willingness to engage in social isolation [3,4]. A limitation of these studies is that they used cross-sectional data; better understanding of the timeline of onset of anxious and depressive symptoms is critical if we are to understand how mental health changes in response to different stages of the pandemic (eg, initial reports, the first national infections, and shelter-in-place).

A handful of studies have employed ecological momentary assessments (EMAs) to assess depression and anxiety more frequently and in near-real-time [5-8]. EMA surveys are sent at predetermined frequencies to the participants' phones as they go about their daily life. This method enables the collection of dense longitudinal data with minimal participant effort relative to in-person research. An initial study in which EMAs from 80 undergraduate students were used to investigate the influence of COVID-19 on mental health and social contact found increased mental health problems but no change in social contact [9].

Our current work combines longitudinal smartphone sensing and EMAs collected from a cohort of Dartmouth College undergraduates to determine the impact on mental health and behaviors during the COVID-19 pandemic. Dartmouth College has four academic terms per year, which are roughly split into 10-week terms followed by 2-week (or longer) breaks. Self-reported mental health has been observed to vary across typical academic terms [8,10]. During the Winter 2020 academic term, Dartmouth College began implementing new policies in response to COVID-19 (Table 1). Beginning on February 4, individuals returning from China were asked to self-isolate for 14 days upon their return to campus. On March 2, the first local COVID-19 case was diagnosed at nearby Dartmouth-Hitchcock Medical Center. On March 10, during the final examination period, the College requested that all students scheduled to leave campus for spring break depart upon completion of their final examinations and further requested that students who were planning to stay on campus during spring break instead depart campus by March 16. Consequently, students were required to rapidly change their travel plans, complete final examinations sooner than expected, or take final examinations online. On March 11, the WHO officially declared COVID-19 a pandemic, at which time all college-sponsored athletic competitions were cancelled. Just one day later, on March 12, the US government implemented policies limiting travel to and from European countries. By the next day, March 13, COVID-19 had been declared a national emergency by the President of the United States. Dartmouth College's spring break started on March 14. On March 16, the college cancelled all gatherings of groups larger than 50 people. Finally, on March 17, the college announced that there would be no in-person, on-campus option to attend classes during the Spring 2020 term. This timeline is particularly relevant because it allows us to identify periods during which we might expect additional changes in mental health and behaviors due to the stress of the pandemic and in light of potential adherence to the "Stay Safe, Stay Home" policies mandated by local and national governments. News coverage in the United States can also serve as a proxy for the perceived severity of the situation, given the rapid transition from a localized outbreak in a country several thousand miles away (China) to an outbreak within a few miles of campus along with a rising number of cases nationally (Figure 1). Increases in anxiety and depression are frequently observed after traumatic events [11-14].

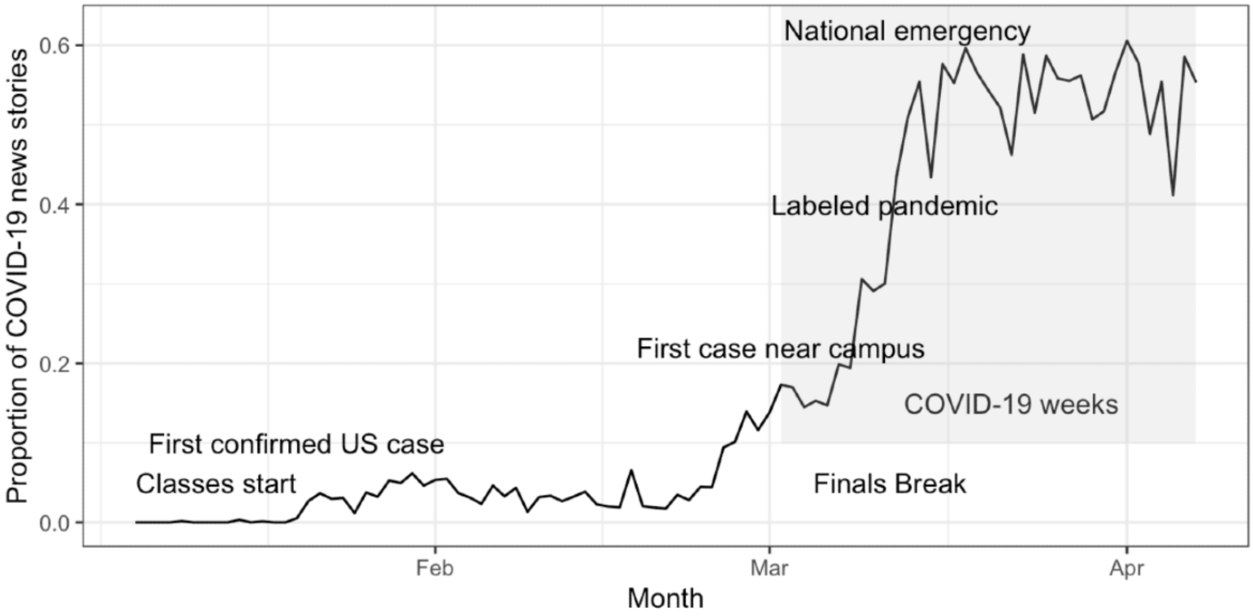
Table 1. Key Dartmouth College academic dates for the Winter 2020 term and relevant COVID-19 events.

Date	Event	Term week
January 6	First day of classes	1
January 20	First confirmed COVID-19[a] case in the United States	3
February 4	Travelers from China asked to self-quarantine	5
March 2	First COVID-19 case near campus	9
March 6	Last day of classes	9
March 9	First day of final examinations	10
March 10	Students asked to leave campus as soon as possible (March 16 at the latest)	10
March 11	WHO[b] labels COVID-19 a pandemic; all Dartmouth College athletics cancelled	10
March 12	Travel between the United States and Europe restricted	10
March 13	COVID-19 declared a national emergency; last day of final examinations	10
March 14	Start of spring break	11 (Break)
March 16	Gatherings of more than 50 individuals cancelled	11 (Break)
March 17	Online-only off-campus learning for Spring 2020 term announced	11 (Break)

[a] COVID-19: coronavirus disease.

[b] WHO: World Health Organization.

Figure 1. Proportion of US news and media stories containing the term *coronavirus* with key COVID-19 and Dartmouth academic term events labeled. Data were obtained from Media Cloud. COVID-19: coronavirus disease.



Objectives

The current work seeks to answer the following questions. First, how and when was the mental health of college students impacted by the COVID-19 epidemic, and were changes in depression and anxiety statistically different from those in previous terms in the same student cohort? Second, how are daily behaviors (as measured by smartphone sensing) and changes in depression and anxiety impacted by COVID-19 media coverage?

Methods

Study Design

All data in the current study were obtained from the second iteration of the StudentLife study [11], which is a longitudinal multimodal study that is designed to follow the experiences of undergraduate students throughout their academic tenure, with a focus on mental health. Study components include smartphone mobile sensing through the StudentLife app [6], EMAs and surveys focusing on a variety of college experience components, and functional neuroimaging [12].

Participants

Data were collected from 219 participants who agreed to provide mobile sensing data via the StudentLife app [6]. One participant was removed from the study because their mobile phone was incompatible with the app, and one participant withdrew within a week of starting the study. Data for both subjects were excluded from further analyses. Of the remaining 217 participants, 147 (67.8%) were female, with an age range for all participants of 18 to 22 years at the time of enrollment. Recruitment for this study began in August 2017 and concluded in November 2018. This study was approved by Dartmouth's Committee for the Protection of Human Subjects.

Academic Terms

At Dartmouth College, the academic calendar consists of a flexible, year-round calendar that is approximately divided into four academic terms, or quarters. Each term consists of 10 weeks, typically followed by 2 (or more) weeks of break. The Winter 2020 academic term (January 6 start date) includes the progression of the COVID-19 pandemic, ranging from the first confirmed case in the United States (Table 1) to the identification of COVID-19 as a global pandemic; these events resulted in policy and regulatory changes at the college, local, and national levels. Academic terms prior to the Winter 2020 term, including the subsequent 2 weeks of break, were included as control terms.

Mobile Sensing and EMAs

Smartphone sensing data and EMA surveys were administered using the StudentLife application (iOS and Android) [6]. The StudentLife app collects data from several of the phone's sensors, including but not limited to the GPS, accelerometer, and lock/unlock status. Anonymized data from the StudentLife app are uploaded to a secure server when a participant is both using WiFi and charging their phone. Data from these sensors are used to assess factors such as the day-to-day and week-by-week impact of workload on the stress, sleep, activity, mood, sociability, mental well-being, and academic performance of the students [6]. Students are prompted weekly by the StudentLife application to complete a few short surveys, administered as EMAs [5]. These EMAs include the Patient Health Questionnaire-4 (PHQ-4), a brief measure of depressive and anxious symptoms [13] that assesses how often individuals were bothered by specific symptoms over the last 2 weeks with values ranging from 0 to 6 for each subscale. The PHQ-4 combines the Patient Health Questionnaire-2 (PHQ-2) and the Generalized Anxiety Disorder-2 (GAD-2). Data coverage was 19.7/24 hours for location, 22.3/24 hours for all other indices, and 80.1% for EMA responses across the duration of the study.

Sedentary Time

Sedentary time, or stationary duration, is computed to measure students' activity or, more precisely, their lack of activity. The app continuously infers physical activities using the Android activity recognition application programming interface [14,15] or iOS Core Motion [16].

Sleep

Sleep was inferred through a combination of passive sensing features (ambient light, movement activity, screen on/off). In this way, 3 features were computed: sleep onset, wake time, and sleep duration. These measures of sleep have been shown to be accurate within 30 minutes for total sleep duration [6].

Location

Density-based spatial clustering of applications with noise (DBSCAN) [17] was used to cluster GPS coordinates to determine the number of locations visited and distance traveled during a given time period. Locations were detected when 3 GPS samples (1 sample every 10 minutes) were within a radius of 30 meters. Distance was calculated in meters traveled between all locations throughout the day.

Phone Usage

Unlock duration is a measurement of time during which a mobile phone is unlocked and the screen is on; it is calculated from the time the user unlocks the phone until they either manually relock the phone or it autolocks due to disuse (the iOS default is 30 seconds, while Android defaults vary by manufacturer; users can also alter this by changing their phone settings). Notification and system services do not influence the measurement of unlock duration. While unlock duration is not an absolute measurement of phone usage, it is the closest approximation implemented in StudentLife. From the start of the study in September 2017 until September 2018, the unlock duration was measured by remotely triggering the mobile phones every 10 minutes, sampling 1 minute every 10 minute period (minimum 10% temporal coverage). If conversation was detected during the 1-minute sampling period, the sampling was extended to 3 min for a maximum of 30% temporal coverage. After September 2018, the mobile phones were remotely triggered every 3 minutes, with subsequent sampling for 1 minute. Lock/unlock behaviors within that minute were recorded in real time, while lock/unlocks for the remaining 2 minutes were logged during the next remote trigger.

COVID-19 News Coverage

To obtain an unbiased measurement of media exposure to COVID-19, the number of news articles published with the term *coronavirus* and the number of all news articles were pulled for the duration of the entire study (August 2017 to 2 weeks after the end of the Winter 2020 academic term). Articles were pulled from a variety of US news outlets, including newspapers and online sources, on the Media Cloud website [18]. The ratio of articles mentioning *coronavirus* to the total number of articles was computed to create a variable indexing the amount of COVID-19 reporting in the media landscape and is presented alongside notable timepoints in the academic term (Figure 1).

Data Processing, Modeling, and Visualization

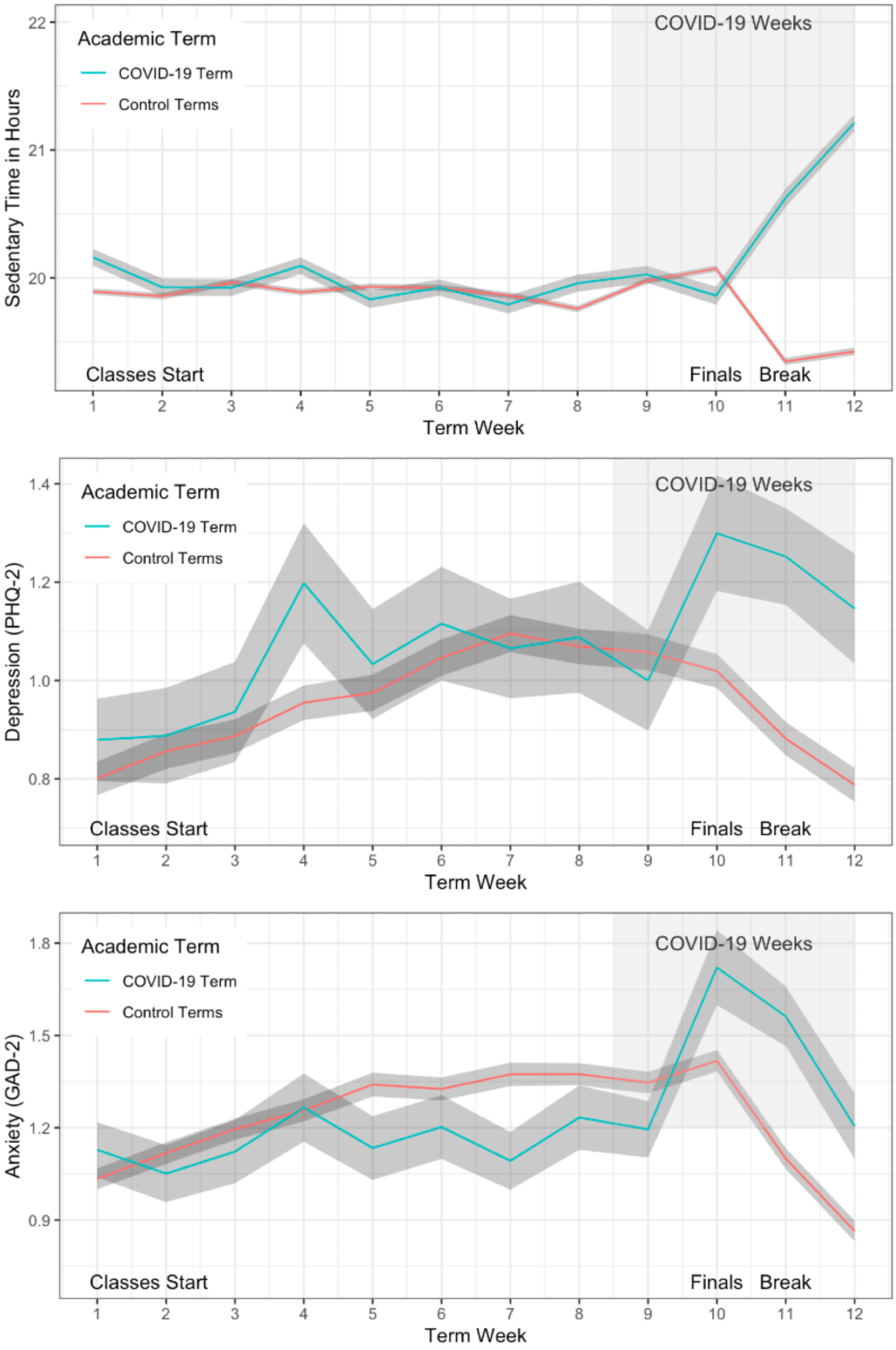
Data processing was performed in R [19] with formatting and development in R Markdown using RStudio [20]. Modeling was implemented in the lme4 [21] and lmerTest [22] packages. Plots were generated using ggplot2 [23]. Result tables were produced using the stargazer package [24]. The objective of the first analysis was to model sedentary time, depression, and anxiety. To observe trends in these domains across the academic term, the mean values of the variables of interest were plotted by week of the term; data were combined from all study participants from all terms except for the Winter 2020 term, which was plotted as a separate line. Standard error was plotted as a shaded ribbon surrounding the mean. Visual representations of the sedentary time and self-reported depression and anxiety can be observed in Figure 2. Linear mixed models fit by log-likelihood were compared to determine if the values were indeed different from prior terms with respect to sedentary time, anxiety, and depression. The terms in each model included a binary factor, “COVIDTerm,” to label whether the term was influenced by COVID-19. The term week was modeled as linear and quadratic factors, along with interactions between the COVIDTerm factor and the term week variables (ie, COVIDTerm×Term Week (Linear)) in subsequent models. Random intercepts were set per subject in all models. Term week variables were scaled to aid model convergence.

Each variable of interest (sedentary time, depression, and anxiety) was individually modeled with COVIDTerm, term week (linear), and random intercepts per subject. The next model added an interaction term between COVIDTerm and term week (Linear). The third model added the term week (quadratic) variable, and the fourth model added an interaction between COVIDTerm and term week (quadratic). For each of the variables of interest, these four models were compared using the analysis of variance function from the base stats package in R. For each variable, the model with the lowest deviance was selected. *P* values were calculated using the Satterthwaite method as implemented in lmerModLmerTest as part of the lmerTest package.

To obtain a daily variable mirroring possible exposure to COVID-19–related news content, a *coronavirus* topic was created on Media Cloud with dates spanning the duration of the study. The proportion of stories including *coronavirus* was downloaded and scaled. Modeling of COVID-19 news was first performed by combining smartphone sensing features and the week of the academic term. Each feature was scaled before being submitted to the model. Each variable, except for the COVIDTerm factor and subject, was scaled to aid model convergence and allow for regression coefficients that can be compared for relative importance. COVID-19 news was inferred with fixed effects of unlock duration, unlock number, sedentary time, sleep duration, number of locations visited, and term week (linear and quadratic) variables, as well as random intercepts for each subject.

To determine if mental health was associated with the COVID-19 news ratio, self-reported depression and anxiety variables were added to the daily model above. Given that individual participants only answered mental health EMAs on a weekly basis, the subsequent models include far fewer time points than the smartphone-sensing only model above; again, this limits the statistical comparability between the models.

Figure 2. Sedentary time (top), depression (middle), and anxiety (bottom) scores across an academic term and the first 2 weeks of break, with the term influenced by the outbreak of the COVID-19 pandemic as a separate line. The shaded ribbons represent the standard error for each week. Weeks influenced by policy changes related to COVID-19 are represented with a shaded box from weeks 9 to 12. Sedentary time was calculated using data from the StudentLife app. Depression and anxiety were measured with the PHQ-2 and GAD-2 scales through the StudentLife app. Control terms include data from the same group of individuals across previous academic terms. COVID-19: coronavirus disease; Finals: final examinations; GAD-2: Generalized Anxiety Disorder-2; PHQ-2: Patient Health Questionnaire-2.



Results

Behavioral and Mental Health Changes Associated With the COVID-19 Pandemic

Compared with all other academic terms, we observed differences in both behavior and mental health between weeks 9 and 12 of the Winter 2020 term, which correspond to the last week of classes, the week of final examinations, and the 2-week spring break. Sedentary time appeared to be very similar across the Winter 2020 and previous terms until week 11 (first week of break), when individuals spent over one hour more per day sedentary compared with a typical term; the sedentary time increased further through the second week of break (Figure 2). Self-reported symptoms of depression and anxiety spiked noticeably in week 10, which corresponds to widespread policy changes at the college, local, and national government levels. Week 10 is 1 week after the first case of COVID-19 was confirmed proximal to Dartmouth College, which occurred on the first day of week 9 of the academic term. Week 10 also corresponds to the time when students were asked to leave campus as soon as possible and when the switch to a remote learning model for the Spring 2020 term was announced. After week 10 in the Winter 2020 term, both depression and anxiety remained consistently elevated above the levels in other terms; however, they decreased at similar rates.

Multiple models for sedentary time, anxiety, and depression were tested (see the Methods section for specific details). For each of these variables, superior fit (measured by lowest deviance) was observed with the most complex model. This model included the COVID-19 term, linear term, and quadratic term week trends as well as the interaction of the COVID-19 term with each of the term week trends; the model also allowed for random intercepts for each participant's data (random effects). Modeling of the academic term affected by COVID-19 compared with academic terms prior to the COVID-19 pandemic identified significantly increased sedentary time, depression, and anxiety (*P*<.001; Table 2). Interactions of the COVID-19 term and quadratic term week regressor for all three variables (*P*<.001) and significant interactions between the COVID-19 term and the linear term week regressor for sedentary time and depression were also observed (*P*<.001 and *P*=.004, respectively).

Table 2. Models of sedentary time, depression, and anxiety by week and by presence of COVID-19 during the academic term.

Variable	Dependent variables					
	Sedentary time (observations=113,864)		Depression (observations=20,323)		Anxiety (observations=113,864)	
	Parameter (SD)	*P* value	Parameter (SD)	*P* value	Parameter (SD)	*P* value
COVID-19[a] term	0.150 (0.008)	<.001	0.176 (0.016)	<.001	0.111 (0.017)	<.001
Term week (linear)	–0.046 (0.003)	<.001	0.016 (0.005)	.004	0.004 (0.006)	.51
Term week (quadratic)	–0.045 (0.003)	<.001	–0.078 (0.006)	<.001	–0.108 (0.006)	<.001
COVID-19 term: term week (linear)	0.138 (0.008)	<.001	0.058 (0.016)	.001	0.079 (0.016)	<.001
COVID-19 term: term week (quadratic)	0.160 (0.008)	<.001	0.064 (0.016)	<.001	0.123 (0.016)	<.001
Constant	–0.040 (0.032)	.21	0.012 (0.046)	.80	0.040 (0.045)	.37

[a] COVID-19: coronavirus disease.

COVID-19 News Coverage, Mental Health, and Mobile Sensing

After establishing broad differences in sedentary time, depression, and anxiety between the Winter 2020 term and previous terms, the next goal was to determine if these behaviors changed in a finer-grained fashion, particularly mirroring the relative news coverage of COVID-19. The proportion of new stories including the term *coronavirus* remained at baseline until early January, with an unsurprisingly large increase appearing at the beginning of March (Figure 1). To ascertain which behaviors changed with increasing proportion of COVID-19 news reports, we included fixed effects for phone usage (unlock duration and unlock number), sedentary time, sleep duration, number of locations visited, and linear and quadratic academic term week regressors. Random intercepts per subject were included in the model. Each variable was scaled to aid convergence of the restricted maximum likelihood model and to obtain regression coefficients that could be compared for relative importance. All variables except sleep duration and distance traveled were significantly associated with the proportion of COVID-19 news reports (*P*<.001, Table 3 [left column], Figure 3 [top]). Phone usage (unlock duration) has the largest positive standardized coefficient, closely followed by the linear term week variable. The number of locations visited has the largest negative standardized coefficient.

In the second model of the COVID-19 news ratio, we again attempted to make inferences with mobile sensing features plus the addition of self-reported anxiety and depression scores. When anxiety and depression were added to the previously used sensing model, we observed that increased anxiety but not depression was significantly associated with a higher COVID-19 news ratio (*P*<.001, Table 3 [right column], Figure 3 [bottom]). The number of phone unlocks (unlock number) is significant in the first model but not in the second model. Since the standardized beta weights are relatively stable across all other variables, it appears that anxiety may have absorbed some of the variance associated with the unlock number in the first model; however, we cannot directly make this comparison due to the different subset of data in the second model. Again, we observed increased phone usage (unlock duration), increased sedentary time, and decreased number of locations visited, with standardized beta weights stable across both models. Both linear and quadratic term weeks were positively associated with the COVID-19 news ratio.

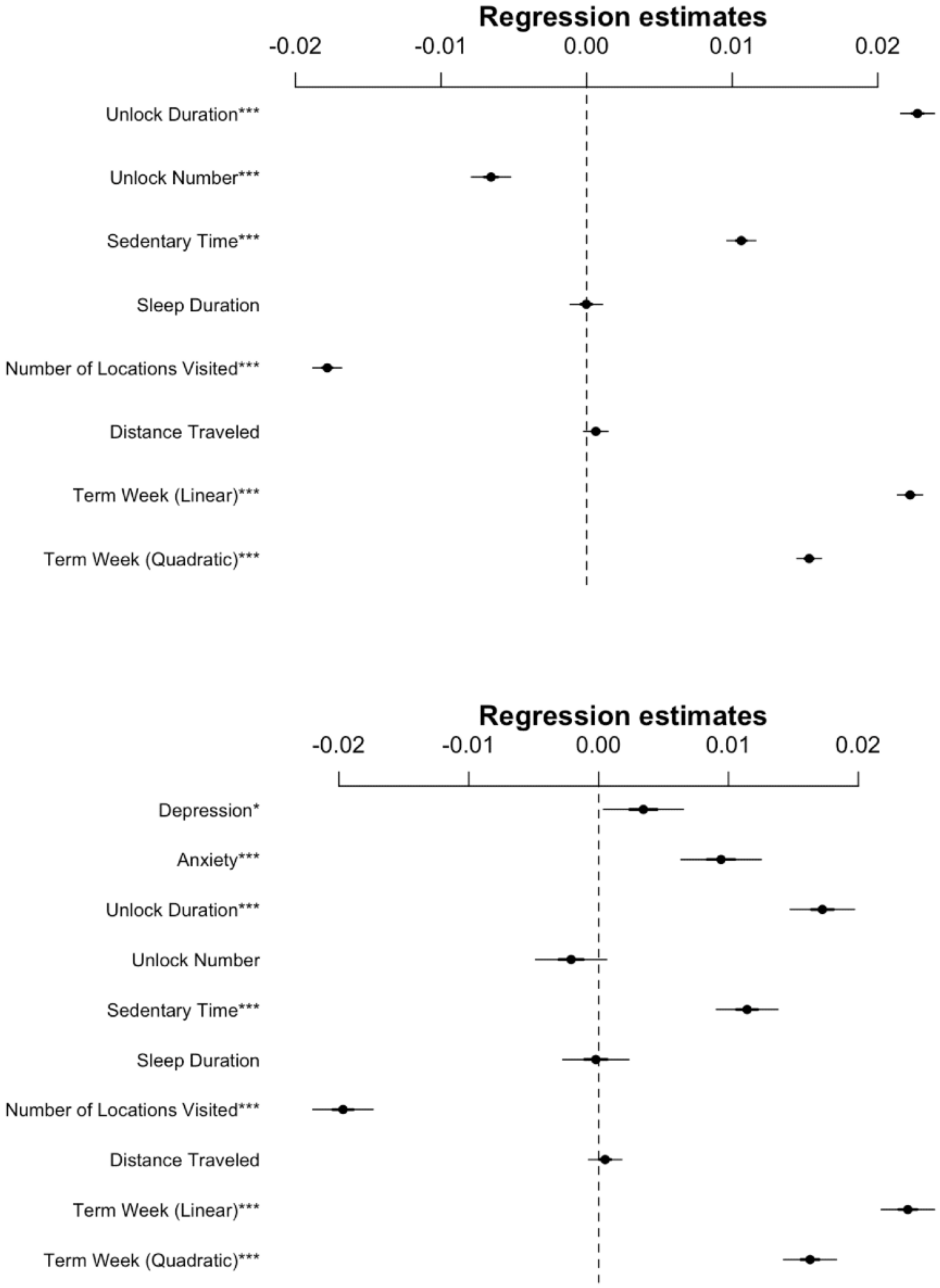
Table 3. Inferences of the proportion of COVID-19 news reports with smartphone features and self-reported mental health variables.

Variable	Dependent variables			
	COVID-19[a] news model 1 (observations=100,300)		COVID-19 news model 2 (observations=18,432)	
	Parameter (SD)	*P* value	Parameter (SD)	*P* value
Depression	N/A[b]	N/A	0.003 (0.002)	.03
Anxiety	N/A	N/A	0.009 (0.002)	<.001
Unlock duration	0.023 (0.001)	<.001	0.017 (0.001)	<.001
Unlock number	–0.007 (0.001)	<.001	–0.002 (0.001)	.13
Sedentary time	0.011 (0.001)	<.001	0.011 (0.001)	<.001
Sleep duration	–0.00002 (0.001)	.97	–0.0002 (0.001)	.87
Number of locations visited	–0.018 (0.001)	<.001	–0.020 (0.001)	<.001
Distance traveled	0.001 (0.0004)	.14	0.001 (0.001)	.44
Term week (linear)	0.022 (0.0004)	<.001	0.024 (0.001)	<.001
Term week (quadratic)	0.015 (0.0004)	<.001	0.016 (0.001)	<.001
Constant	0.029 (0.002)	<.001	0.028 (0.002)	<.001

[a] COVID-19: coronavirus disease.

[b] Not applicable.

Figure 3. Coefficient plots from mixed linear models of COVID-19 news inferred with mobile smartphone features (top) or mobile smartphone features and self-reported mental health (bottom). Intercept and random intercepts per subject are not plotted. *, **, and ***: *P*<.05, <.01, and <.001, respectively. Where applicable, exact *P* values are shown in Table 3.



Discussion

Principal Findings

During the outbreak of a pandemic, the primary focus is on the pathogen and its influences on physical health. Mental health and behavioral changes are often considered secondary concerns. In the present study, we leveraged longitudinal data in a cohort of 217 college students to demonstrate the impact of the COVID-19 pandemic on mental health and behavior compared to previous academic terms. During the Winter 2020 academic term, sedentary time increased along with symptoms of anxiety and depression. Subsequent analyses examined the association between increased COVID-19 news coverage and behavior (inferred from mobile sensing data) and mental health. We found that with increasing COVID-19–related news, individuals were more sedentary, visited fewer locations (as inferred from GPS tracking), and showed increases in anxiety and depression. These analyses identified behavioral changes from smartphone sensing that are consistent with individuals adhering to the "Stay Safe, Stay Home" policies implemented by local and national governments.

During the Winter 2020 academic term, increased depression, anxiety, and sedentary time were observed, suggesting a large mental health and behavioral impact beyond the actual reach of the SARS-CoV-2 pathogen. The days and weeks leading up to and including the final examination period are a particularly stressful time for students [10]. We typically find that participants report relatively high scores of depression and anxiety during this time; however, these scores recover to baseline over the break [8]. The cyclical nature of college students' mental health within a typical academic term provides a unique control in our study that might be otherwise difficult to untangle from periodic rises and falls in stress and anxiety during March as colleges across the United States conduct final examinations. Understanding behaviors during typical academic cycles can be fruitful for determining how everyday events impact students' mental health and behaviors. Rare events such as the COVID-19 pandemic provide a unique opportunity to examine how mental health and behaviors deviate from baseline. To account for the cyclical nature of mental health throughout the academic term, we used linear and quadratic week of academic term variables, then looked for interactions and the main effects of the most recent (COVID-19–related) term. The increased depression, anxiety, and sedentary time, which were above and beyond what would normally be observed during a typical term, were attributed to the COVID-19 pandemic. Additionally, we did not observe a return to baseline for any of these 3 variables over the break, which stands in stark contrast to our previous study of this same cohort in prior academic terms. We did observe decreases in anxiety and depression that paralleled the typical post–final examination drop, which suggests some resilience in the face of COVID-19; however, the overall values remained elevated above those observed during typical academic breaks. This may also be consistent with the adjustment period to the rapidly changing landscape of social media, policies, and media coverage [25].

Individuals are generally more active over break periods; however, we observed a large increase in sedentary behavior, which can be attributed to the COVID-19 pandemic. Studies show that physical activity, particularly aerobic activity, reduces self-reported depressive symptoms with a similar efficacy to that of low-dose antidepressants [26]. Other studies have shown that increased sedentary time along with increased phone usage is implicated in depression and anxiety [12,27,28]. Taken as a whole, these findings suggest that during stay-at-home orders, individuals should increase physical activity and limit screen time in an attempt to lessen depressive symptoms.

As COVID-19 news reporting intensified, we observed increased sedentary behavior and duration of phone usage along with decreases in the number of locations visited and a decreased number of phone unlocks. Initially, the number of phone unlocks was somewhat surprisingly inversely associated with COVID-19 news; however, this is likely due to increased phone unlock duration (ie, screen time), which would otherwise span multiple phone unlocks. The decreased number of locations visited is consistent with the "Stay Home, Stay Safe" polices many governments have implemented; it also stands in contrast to initial work on social distancing among college students in the Netherlands, where social distancing policies were not implemented during the examined time period [9]. In the combined sensing and mental health model, the addition of depression and anxiety to the sensing model showed a strong inferential link between anxiety and COVID-19 news, while depression was marginally significant. In the combined model, the number of unlocks, sleep duration, and distance traveled were again not significant.

Primary takeaways from models inferring the proportion of COVID-19 news stories suggest that during the start of the COVID-19 pandemic, students were more depressed and anxious, used their phones more, visited fewer locations, and spent more time sedentary. At this critical time of increased depression and anxiety, we issue a call to public health officials and individual citizens to raise public awareness about the benefits of aerobic exercise and unplugging from technology (moderating phone usage), as each of these has previously been shown to have positive effects on alleviating anxiety and depression [8,26,28,29]. COVID-19 arrived locally during week 9 of the academic term. By the onset of the COVID-19 pandemic (academic term week 10), a significant deterioration in mental health and multiple behavioral changes were observed, which synchronizes with the implementation of rapid policy changes at the college, local, and national levels. These findings suggest a greatly expanded scope of the impact of the COVID-19 pandemic beyond the illness and deaths directly associated with the SARS-CoV-2 pathogen.

Limitations and Future Directions

The current study has a variety of limitations, although most provide incentives for future research. First, our participants are approximately the same age, are undergraduate college students with smartphones compatible with the StudentLife app, and are willing to participate in a multiyear longitudinal research study; this limits the generalizability of the current findings to the general population. Simultaneously, it provides distinct advantages, such as longitudinal measurement across behavioral cycles, wherein previous academic terms can be compared to the term affected by COVID-19. A second limitation is the moderate number of individuals included in the study, with 217 total participants across all terms and 178 across the COVID-19 term (83%). Despite the moderate sample size, strong significant effects on mental health and behavior were observed, suggesting robust effects.

While smartphone sensing is quite robust, there are some limitations to the interpretation of the available data. When mobility is decreased, such as during a stay-at-home order, individuals may not have their mobile phones with them at all times, which could lead to overestimation of sedentary time. Additionally, participants may be preferentially accessing larger screens (eg, tablets or laptops); therefore, phone usage (as measured by screen unlock duration or number of unlocks) may underestimate the total amount of screen time. Even so, we observed increased phone usage despite the possibility of underestimated changes in the total amount of screen time. Future work should also identify the types of screens used during this increased consumption period and quantify the relative amounts of news, social media, and other content consumed. In further work, we could also use smartwatches to improve the measurement of behaviors such as sedentary time and allow for more frequent sampling of phone usage, location, and other measures.

While this work primarily focuses on the initial days of the COVID-19 pandemic, future work would be well suited to investigating differences in mental health and behaviors between typical residential academic terms and terms that have been shifted from residential to online coursework due to COVID-19. Furthermore, identifying causal patterns between the pandemic, policy changes on national, local, and college levels, mental health, and behavior could provide further insight into the optimal development of interventions designed to mitigate mental health crises in the face of global crises. These findings suggest a greatly expanded scope of the impact of the COVID-19 pandemic beyond that directly associated with the SARS-CoV-2 pathogen.

Conclusions

Understanding behaviors during typical academic cycles can be fruitful for determining how everyday events impact students' mental health and behaviors. Fringe events such as the COVID-19 pandemic provide opportunities to examine how mental health and behaviors deviate from baseline.

This study provides preliminary insight into mental health and related behaviors during the initial phases of the COVID-19 pandemic. Depression, anxiety, and sedentary time increased as the COVID-19 pandemic encroached on a college campus in parallel with large-scale policy changes. Using a mixed linear model of smartphone mobile sensing and self-reported mental health questions, we were able to infer the proportion of COVID-19–related news stories; moreover, we could validate that participants' mental health and related behaviors changed in lockstep with increased media coverage and proximity of the pandemic. For these college students, the early days of the pandemic coincided with what is typically a time of increased time and depression, and we observed altered behavioral patterns and a decrease in mental health above and beyond typical academic terms. Much more work remains to be done to understand how behaviors and mental health change and interact in the face of monumental adverse global events.

Acknowledgments

This work was supported by NIMH 5R01MH059282-14.

Conflicts of Interest

None declared.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
2. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
3. Liu X, Liu J, Zhong X. Psychological State of College Students During COVID-19 Epidemic. SSRN Journal 2020 Mar 04:Preprint [FREE Full text] [doi: 10.2139/ssrn.3552814]
4. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
5. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
6. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.: ACM Press; 2014 Presented at: 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, WA p. 3-14. [doi: 10.1145/2632048.2632054]
7. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
8. Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-26. [doi: 10.1145/3191775]
9. Fried E, Papanikolaou F, Epskamp S. Mental Health and Social Contact During the COVID-19 Pandemic: An Ecological Momentary Assessment Study. PsyArxiv 2020 Apr 24:28. [doi: 10.31234/osf.io/36xkp]

10. daSilva AW, Huckins JF, Wang R, Wang W, Wagner DD, Campbell AT. Correlates of Stress in the College Environment Uncovered by the Application of Penalized Generalized Estimating Equations to Mobile Sensing Data. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e12084 [FREE Full text] [doi: 10.2196/12084] [Medline: 30888327]
11. StudentLife Study. URL: https://studentlife.cs.dartmouth.edu/ [accessed 2020-06-11]
12. Huckins JF, daSilva AW, Wang R, Wang W, Hedlund EL, Murphy EI, et al. Fusing Mobile Phone Sensing and Brain Imaging to Assess Depression in College Students. Front Neurosci 2019;13:248 [FREE Full text] [doi: 10.3389/fnins.2019.00248] [Medline: 30949024]
13. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
14. Google APIs for Android. ActivityRecognitionApi URL: https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi [accessed 2020-06-11]
15. Wang R, Aung M, Abdullah S, Brian R, Campbell A, Choudhury T, et al. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12-16, 2016; Heidelberg, Germany p. 886-897. [doi: 10.1145/2971648.2971740]
16. Apple Developer. 2017. iOS Core Motion URL: https://developer.apple.com/reference/coremotion [accessed 2020-06-11]
17. Daszykowski M, Walczak B. Density-Based Clustering Methods. In: Comprehensive Chemometrics. Amsterdam: Elsevier; 2009:654.
18. Media Cloud. URL: https://mediacloud.org/ [accessed 2020-06-11]
19. R Core Team. R-Project. Vienna, Austria: R Foundation for Statistical Computing; 2018. R: A Language and Environment for Statistical Computing URL: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2020-06-11]
20. RStudio. URL: https://rstudio.com/ [accessed 2020-06-11]
21. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
22. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
23. tidyverse. ggplot2 URL: https://ggplot2.tidyverse.org [accessed 2020-06-11]
24. Hlavac M. R-Project. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI); 2018. stargazer: Well-Formatted Regression and Summary Statistics Tables URL: https://cran.r-project.org/package=stargazer [accessed 2020-06-11]
25. Jacobson N, Lekkas D, Price G, Heinz M, Song M, O'Malley AJ, et al. Flattening the Mental Health Curve: COVID-19 Stay-at-Home Orders Are Associated With Alterations in Mental Health Search Behavior in the United States. JMIR Ment Health 2020 Jun 01;7(6):e19347 [FREE Full text] [doi: 10.2196/19347] [Medline: 32459186]
26. Craft L, Landers D. The Effect of Exercise on Clinical Depression and Depression Resulting from Mental Illness: A Meta-Analysis. J Sport Exercise Psy 1998 Dec;20(4):339-357. [doi: 10.1123/jsep.20.4.339]
27. Zhai L, Zhang Y, Zhang D. Sedentary behaviour and the risk of depression: a meta-analysis. Br J Sports Med 2015 Jun;49(11):705-709. [doi: 10.1136/bjsports-2014-093613] [Medline: 25183627]
28. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017 Jan 01;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]
29. Twenge JM, Joiner TE, Rogers ML, Martin GN. Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time. Clin Psychol Sci 2017 Nov 14;6(1):3-17. [doi: 10.1177/2167702617723376]

Abbreviations

COVID-19: coronavirus disease
DBSCAN: density-based spatial clustering of applications with noise
EMA: ecological momentary assessment
GAD-2: Generalized Anxiety Disorder-2
PHQ-2: Patient Health Questionnaire-2
PHQ-4: Patient Health Questionnaire-4
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 12.05.20; peer-reviewed by R Cruz Martínez, R Ho, D Leightley, K Wall; comments to author 28.05.20; revised version received 06.06.20; accepted 09.06.20; published 17.06.20.

Please cite as:
Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, Wu J, Obuchi M, Murphy EI, Meyer ML, Wagner DD, Holtzheimer PE, Campbell AT
Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study
J Med Internet Res 2020;22(6):e20185
URL: http://www.jmir.org/2020/6/e20185/
doi:10.2196/20185
PMID:32519963

Short Paper

Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis

Jean Louis Pépin[1,2], MD, PhD; Rosa Maria Bruno[3,4], MD; Rui-Yi Yang[5], PhD; Vincent Vercamer[5], PhD; Paul Jouhaud[5], MEng; Pierre Escourrou[6,7], MD, PhD; Pierre Boutouyrie[3,4], MD, PhD

[1]HP2 (Hypoxia and Physio-Pathologies) Laboratory, Inserm (French National Institute of Health and Medical Research) U1042, University Grenoble Alpes, Grenoble, France

[2]EFCR (Cardiovascular and Respiratory Function) Laboratory, Grenoble Alpes University Hospital, Grenoble, France

[3]Inserm (French National Institute of Health and Medical Research) U970, Hôpital Européen Georges Pompidou, Assistance Publique Hôpitaux de Paris, Paris, France

[4]Université de Paris, Paris, France

[5]Withings, Issy les Moulineaux, France

[6]Hôpital Béclère, Assistance Publique Hôpitaux de Paris, Paris, France

[7]Université Paris Saclay, Paris, France

Corresponding Author:
Jean Louis Pépin, MD, PhD
HP2 (Hypoxia and Physio-Pathologies) Laboratory
Inserm (French National Institute of Health and Medical Research) U1042
University Grenoble Alpes
CHU (University Hospital) Michallon
Grenoble, 38043
France
Phone: 33 476768473
Email: JPepin@chu-grenoble.fr

Abstract

Background: In the context of home confinement during the coronavirus disease (COVID-19) pandemic, objective, real-time data are needed to assess populations’ adherence to home confinement to adapt policies and control measures accordingly.

Objective: The aim of this study was to determine whether wearable activity trackers could provide information regarding users' adherence to home confinement policies because of their capacity for seamless and continuous monitoring of individuals’ natural activity patterns regardless of their location.

Methods: We analyzed big data from individuals using activity trackers (Withings) that count the wearer’s average daily number of steps in a number of representative nations that adopted different modalities of restriction of citizens’ activities.

Results: Data on the number of steps per day from over 740,000 individuals around the world were analyzed. We demonstrate the physical activity patterns in several representative countries with total, partial, or no home confinement. The decrease in steps per day in regions with strict total home confinement ranged from 25% to 54%. Partial lockdown (characterized by social distancing measures such as school closures, bar and restaurant closures, and cancellation of public meetings but without strict home confinement) does not appear to have a significant impact on people’s activity compared to the pre-pandemic period. The absolute level of physical activity under total home confinement in European countries is around twofold that in China. In some countries, such as France and Spain, physical activity started to gradually decrease even before official commitment to lockdown as a result of initial less stringent restriction orders or self-quarantine. However, physical activity began to increase again in the last 2 weeks, suggesting a decrease in compliance with confinement orders.

Conclusions: Aggregate analysis of activity tracker data with the potential for daily updates can provide information regarding adherence to home confinement policies.

(J Med Internet Res 2020;22(6):e19787) doi:10.2196/19787

KEYWORDS

wearable activity trackers; pandemic; COVID-19; home confinement; lockdown; monitoring; wearables; tracking

Introduction

Nationwide total home confinement is the most significant measure that has been taken to prevent the spread of coronavirus disease (COVID-19) infection [1]. Currently, over 3 billion people worldwide, one-quarter of the world's population, are confined to their homes. However, the timing and stringency of governmental decisions have been heterogeneous; some governments have imposed total lockdown, while others have required partial or no confinement. Objective, real-time measures to assess populations' adherence to confinement are essential to adapt policies and control measures accordingly.

We asked whether wearable activity trackers could provide this information because of their capacity for seamless and continuous monitoring of individuals' natural activity patterns regardless of their location [2]. Data from activity trackers enable the compilation of synchronized big data resources on human behavior with high geographical and temporal resolution. In the context of the COVID-19 pandemic, activity trackers provide a valuable data set that objectively documents the time course of adherence to home confinement worldwide in response to the outbreak.

Methods

We analyzed data from approximately 742,000 individuals using activity trackers that count the wearer's average daily number of steps (Withings) in a number of representative nations that adopted different modalities of restriction of citizens' activities. We selected the nations according to a compromise between exposure to the COVID-19 pandemic, rules of lockdown, and availability of data from a large number of activity tracker users.

The pre-pandemic period was used as a reference. For each individual, we calculated the average daily number of steps between December 1, 2019, and the date of lockdown, representing the pre-pandemic period, and the average daily number of steps between the date of lockdown and the analysis time point (4 weeks for China). These data were then aggregated across countries or regions. The paired Wilcoxon nonparametric test was used to compare the number of steps per day before and during lockdown. The data presented extend to April 13, 2020, and are averaged by days and over countries or provinces/states.

The activity tracker used in this study was a wristwatch with an embedded accelerometer that counts steps. The performance of this tracker is reported to be one of the best among available devices [3]. The activity tracker provides the most accurate measures of step count under all three important physiological conditions (ie, treadmill, over-ground, and 24-hour conditions). The same accelerometer and algorithm were used for all individuals included in the analysis. All activity tracker wearers were informed that the anonymized data collected from them could be used for research purposes, and they provided informed consent before starting to use the activity tracker. They were allowed to withdraw their consent at any time and request deletion of their individual data.

Results

Table 1 shows the number of users per country or province/state, percentage of female users, lockdown initiation dates, rules, and percentage of decrease in steps during lockdown. The most demonstrative countries are presented in Figure 1, illustrating the time course of the step count and enhancing the data in Table 1. Before the epidemic, all countries showed a stable mean number of steps per day, with periodic and reproducible decreases during weekends. In countries adopting a total lockdown, a marked decrease (from 25% to 54%) in the number of steps following the official dates of home confinement can be clearly identified (Figure 1A). Partial lockdown (characterized by social distancing measures such as school closures, bar and restaurant closures, and cancellation of public meetings but without strict home confinement) does not appear to have a significant clinical impact on people's activity compared to the pre-pandemic period (Figure 1B), with similar activity patterns to those in nations without any restriction orders (Figure 1C).

Table 1. Characteristics of the studied population.

Country	Province or state	Size of user population[a]	Proportion of women (%)[a]	Mean age (years)	Lockdown rules	Lockdown date	Baseline steps per day	Lockdown steps per day	Decrease in steps (%)	*P* value
Australia	N/A[b]	10,000	42	42	Partial	2020-03-23	5765	5302	8.0	<.001
Canada	N/A	10,000	38	43	None	N/A	5049	4708	6.8	<.001
China	N/A	10,000	19	36	Total	2020-01-23	4108	3034	26.1	<.001
China	Hubei	100	14	35	Total	2020-01-23	4375	1943	55.6	<.001
France	N/A	100,000	43	43	Total	2020-03-17	4604	3342	27.4	<.001
Germany	N/A	100,000	37	46	Partial	2020-03-16	5349	5416	–1.3	<.001
Ireland	N/A	10,000	38	42	Total	2020-03-28	5326	5356	–0.6	<.001
Italy	N/A	10,000	31	45	Total	2020-03-10	5445	3918	28.0	<.001
Italy	Lodi	100	29	45	Total	2020-02-21	5640	5035	10.7	<.001
Japan	N/A	100,000	29	43	Total	2020-04-07	5460	4581	16.1	<.001
Netherlands	N/A	10,000	38	44	None	N/A	5193	5180	0.3	<.001
Singapore	N/A	1000	33	41	None	N/A	6127	5860	4.3	<.001
Spain	N/A	10,000	36	46	Total	2020-03-15	6215	3638	41.5	<.001
Sweden	N/A	10,000	34	44	None	N/A	5681	6004	–5.7	<.001
Switzerland	N/A	10,000	40	44	Partial	2020-03-16	5325	4947	7.1	<.001
United Kingdom	N/A	100,000	39	43	Total	2020-03-23	5690	5249	7.8	<.001
United States	N/A	100,000	43	43	Partial	2020-03-22	5287	4912	7.1	<.001
United States	California	100,000	38	43	Total	2020-03-19	5508	5013	9.0	<.001
United States	Florida	10,000	44	46	Partial	2020-03-17	5303	5225	1.5	<.001
United States	Illinois	10,000	41	42	Total	2020-03-21	5415	4966	8.3	<.001
United States	New Jersey	10,000	38	43	Total	2020-03-21	5297	4693	11.4	<.001
United States	Pennsylvania	10,000	43	44	Partial	2020-03-19	5186	4974	4.1	<.001
United States	New York	10,000	39	42	Partial	2020-03-22	5776	4499	22.1	<.001
United States	Nevada	1000	42	45	Partial	2020-03-21	5391	4902	9.1	<.001

[a] The number of users having activity data on a given day is subject to variation; the numbers given in the table are representative orders of magnitude of the daily number of users having activity data.

[b] Not applicable.

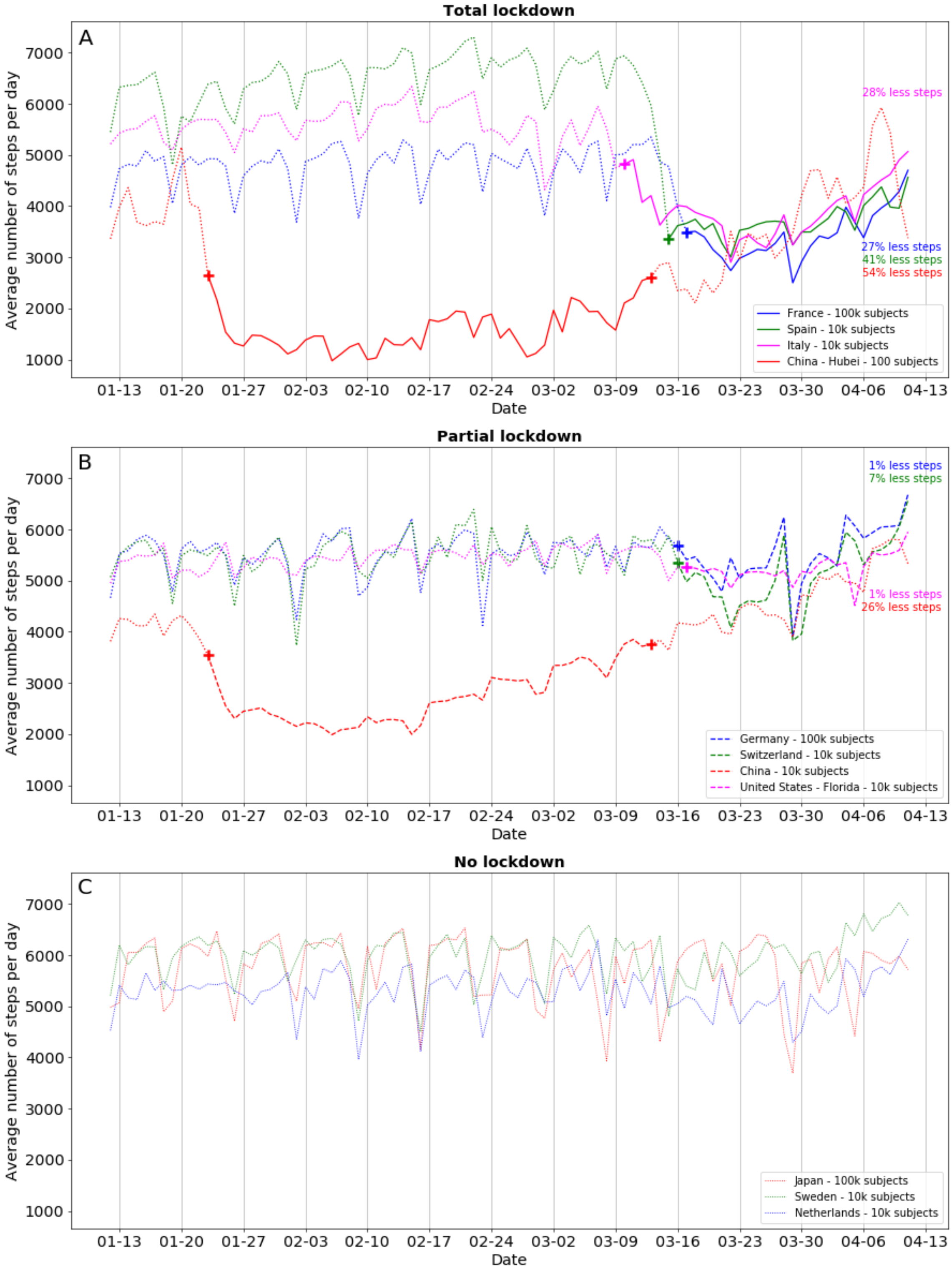


Figure 1. Trajectories of average daily number of steps recorded by activity trackers from January 13, 2020, to April 13, 2020, in a number of representative countries worldwide adopting total (A), partial (B), or no (C) lockdown. Solid lines indicate total lockdown periods, dashed lines indicate partial lockdown periods, and dotted lines indicate no lockdown. Crosses indicate the start and end dates of lockdown in the different countries.

Discussion

Principal Findings

The absolute level of physical activity under total home confinement in European countries is around twofold that in China (Table 1); this may be due to stricter governmental rules in China or different national temperaments. Interestingly, in some countries, such as France and Spain, physical activity started to gradually decrease even before official commitment to lockdown as a result of initial less stringent restriction orders or self-quarantine. However, physical activity began to increase again in the last 2 weeks, suggesting a decrease in compliance with confinement orders. Countries with partial or no lockdown policies had marginal or no changes in walking habits. Ireland was the only country with enforced confinement that showed no change in step counts. Regarding the magnitude and significance of the changes, all differences were highly significant in statistical terms (*P*<.001) because of the large number of users and because of consistent trends among users (the vast majority of users changed their step counts in the same direction, even for small changes). Clinical significance is thus arbitrary, and the weekly trends suggest spontaneous group changes (decreases on weekends). In fully locked-down countries, with the exception of Ireland, the number of steps decreased below the maximum on weekends; this shows overall good compliance with lockdown rules.

Conclusion

Aggregate analysis of activity tracker data, with the potential for daily updates, can inform governments and stakeholders on adherence to home confinement policies as well as their efficacy without violating citizens' privacy [4]. The data allow comparison of the effectiveness of different government policies. Finally, quantification of physical activity patterns, particularly leisure versus occupational patterns, and their consequences on cardiometabolic health are important because sport and leisure physical activities have been shown to have positive effects on cardiometabolic health, whereas occupational physical activity has not [5]. Data emerging from studies conducted during lockdown will help to address this issue.

Acknowledgments

We thank Alison Foote, Grenoble Alpes University Hospital, France, for critically editing the manuscript. JLP is supported by a research grant from the French National Research Agency (ANR-12-TECS-0010) in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) and the "e-Health and Integrated Care" and "Trajectories Medicine" Chairs of Excellence at Université Grenoble Alpes (MIAI Grenoble Alpes, ANR-19-P3IA-0003).

Authors' Contributions

JLP and PB conceived the study; RY, VV, and PJ acquired and analyzed the data; and JLP, RMB, PE, and PB drafted and significantly revised the manuscript.

Conflicts of Interest

RY, VV, and PJ are employees of Withings, a manufacturer of wearable activity tracking systems; PE is a remunerated scientific advisor to Withings; and JLP and PB are informal unpaid scientific advisors to Withings.

References

1. Wang G, Zhang Y, Zhao J, Zhang J, Jiang F. Mitigate the effects of home confinement on children during the COVID-19 outbreak. Lancet 2020 Mar;395(10228):945-947. [doi: 10.1016/s0140-6736(20)30547-x] [Medline: 32145186]
2. Frie K, Hartmann-Boyce J, Jebb S, Oke J, Aveyard P. Patterns in Weight and Physical Activity Tracking Data Preceding a Stop in Weight Monitoring: Observational Analysis. J Med Internet Res 2020 Mar 17;22(3):e15790 [FREE Full text] [doi: 10.2196/15790] [Medline: 32181749]
3. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
4. Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, et al. Aggregated mobility data could help fight COVID-19. Science 2020 Apr 10;368(6487):145-146. [doi: 10.1126/science.abb8021] [Medline: 32205458]
5. Climie RE, Boutouyrie P, Perier M, Chaussade E, Plichart M, Offredo L, et al. Association Between Occupational, Sport, and Leisure Related Physical Activity and Baroreflex Sensitivity. Hypertension 2019 Dec;74(6):1476-1483. [doi: 10.1161/hypertensionaha.119.13461] [Medline: 31679418]

Abbreviations

COVID-19: coronavirus disease

Edited by I Tagliente, G Eysenbach; submitted 01.05.20; peer-reviewed by R Cruz Martínez, D Filos, M Fahimi; comments to author 22.05.20; revised version received 03.06.20; accepted 04.06.20; published 19.06.20.

Please cite as:
Pépin JL, Bruno RM, Yang RY, Vercamer V, Jouhaud P, Escourrou P, Boutouyrie P
Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis
J Med Internet Res 2020;22(6):e19787
URL: http://www.jmir.org/2020/6/e19787/
doi:10.2196/19787
PMID:32501803

Original Paper

A Double Triage and Telemedicine Protocol to Optimize Infection Control in an Emergency Department in Taiwan During the COVID-19 Pandemic: Retrospective Feasibility Study

Chien-Hao Lin[1], MD; Wen-Pin Tseng[1], MD; Jhong-Lin Wu[1], MD; Joyce Tay[1], MD; Ming-Tai Cheng[1], MD, MPH; Hooi-Nee Ong[1], MD; Hao-Yang Lin[1], MD; Yi-Ying Chen[1], MD; Chih-Hsien Wu[1], MD; Jiun-Wei Chen[1], MD; Shey-Ying Chen[1], MD, PhD; Chang-Chuan Chan[2], SCD; Chien-Hua Huang[1*], MD, PhD; Shyr-Chyr Chen[1*], MD, MBA

[1]Department of Emergency Medicine, National Taiwan University Hospital, National Taiwan University College of Medicine, Taipei, Taiwan

[2]Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chien-Hua Huang, MD, PhD
Department of Emergency Medicine
National Taiwan University Hospital
National Taiwan University College of Medicine
No 7, Zhongshan S Road
Zhongzheng District
Taipei, 100
Taiwan
Phone: 886 2 233123456 ext 62831
Email: chhuang5940@ntu.edu.tw

Abstract

Background: Frontline health care workers, including physicians, are at high risk of contracting coronavirus disease (COVID-19) owing to their exposure to patients suspected of having COVID-19.

Objective: The aim of this study was to evaluate the benefits and feasibility of a double triage and telemedicine protocol in improving infection control in the emergency department (ED).

Methods: In this retrospective study, we recruited patients aged ≥20 years referred to the ED of the National Taiwan University Hospital between March 1 and April 30, 2020. A double triage and telemedicine protocol was developed to triage suggested COVID-19 cases and minimize health workers' exposure to this disease. We categorized patients attending video interviews into a telemedicine group and patients experiencing face-to-face interviews into a conventional group. A questionnaire was used to assess how patients perceived the quality of the interviews and their communication with physicians as well as perceptions of stress, discrimination, and privacy. Each question was evaluated using a 5-point Likert scale. Physicians' total exposure time and total evaluation time were treated as primary outcomes, and the mean scores of the questions were treated as secondary outcomes.

Results: The final sample included 198 patients, including 93 cases (47.0%) in the telemedicine group and 105 cases (53.0%) in the conventional group. The total exposure time in the telemedicine group was significantly shorter than that in the conventional group (4.7 minutes vs 8.9 minutes, *P*<.001), whereas the total evaluation time in the telemedicine group was significantly longer than that in the conventional group (12.2 minutes vs 8.9 minutes, *P*<.001). After controlling for potential confounders, the total exposure time in the telemedicine group was 4.6 minutes shorter than that in the conventional group (95% CI −5.7 to −3.5, *P*<.001), whereas the total evaluation time in the telemedicine group was 2.8 minutes longer than that in the conventional group (95% CI −1.6 to −4.0, *P*<.001). The mean scores of the patient questionnaire were high in both groups (4.5/5 to 4.7/5 points).

Conclusions: The implementation of the double triage and telemedicine protocol in the ED during the COVID-19 pandemic has high potential to improve infection control.

(J Med Internet Res 2020;22(6):e20586) doi:10.2196/20586

KEYWORDS

COVID-19; triage; emergency department; health care workers; infection control; telemedicine

Introduction

Background

Since the beginning of the 20th century, various infectious diseases have repeatedly threatened both population health and health care systems worldwide. The rapid growth of international transportation has paved the way for the transmission of infectious diseases across regions, including severe acute respiratory syndrome (SARS), H1N1 influenza, Middle Eastern respiratory syndrome (MERS), and Ebola virus [1-4]. In 2020, coronavirus disease (COVID-19) was first reported in Wuhan, China; this virus spread globally through large-scale transmission and continues to pose great challenges to medical, public health, and socioeconomic systems worldwide [5]. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), identified as the causative pathogen of COVID-19, is highly contagious [6]. It is primarily transmitted through droplets and close contact, even in the early course of the disease, as well as from asymptomatic patients [6]. The response strategies for controlling COVID-19 outbreaks include early diagnosis, patient isolation, symptomatic monitoring, and quarantine [7].

During the response to the COVID-19 pandemic, a considerable number of frontline health care workers have been infected with the new coronavirus. In China, the first COVID-19 case among health care workers was reported on January 20, 2020; as of February 11, 2020, health care workers represented 1716/44,672 (3.8%) of all patients with laboratory-confirmed COVID-19 and 247/1668 (14.8%) of critical cases, with 5 deaths [8]. In Italy, health care workers accounted for 2026/22,512 (9.0%) of confirmed cases as of March 15, 2020 [9]. According to a recent report from the US Centers for Disease Control and Prevention (CDC), this figure in the United States was 9282/315,531 (2.9%); 2.1%-4.9% of health care workers required intensive care unit admission, and 27 deaths were reported [10]. By April 8, 2020, COVID-19 had affected 22,073 health care workers in 52 countries [11]. The transmission of SARS-CoV-2 among frontline health care workers may result from their long, direct exposure to many infected patients and from a lack of personal protective equipment (PPE) [12]. Therefore, further infection control measures should be undertaken to minimize the direct exposure of health care workers to patients suspected of having COVID-19 [7].

Importance

Emergency departments (EDs) are at the frontline of the health care response to the COVID-19 pandemic and are responsible for rapid and safe triage of patients and isolation of suspected patients with COVID-19 from patients with noninfectious diseases. Therefore, establishing a screening protocol for patients with suspected COVID-19 and isolating them in a separate, well-ventilated space for clinical interviews can facilitate infection control [7]. Moreover, due to the shortages of PPE worldwide, its use should be optimized and prioritized based on critical requirements at health care facilities. Additional strategies to minimize exposure risk, such as implementing alternatives to face-to-face triage and hospital visits, can not only prevent health care workers from being exposed to COVID-19 but can also reduce PPE use. The combination of a specialized triage approach and telemedicine may be an alternative to reach this goal.

Telemedicine can be defined broadly as the use of telecommunications technology to provide medical information and services [13]. With the evolution in technology and online services, telemedicine may prove a compelling alternative to conventional acute, chronic, and preventive care [14-17]. Telemedicine, particularly video consultations, has been promoted to reduce the risk of disease transmission [18-23]. Certain reports have recently described potential benefits and applications of telemedicine during the COVID-19 pandemic [24-27]. However, the effectiveness of telemedicine in improving infection control has not been well studied, as most evidence has been obtained from clinical cases, and some studies merely propose theoretical assumptions.

Goal of This Study

As of May 2, 2020, Taiwan had 432 confirmed COVID-19 cases, and a total of 63,713 patients were tested. Nearly 97% of the Taiwanese population have access to the health care system and are covered by the National Health Insurance; hence, they can readily access health care [28]. EDs were the first units that responded to the COVID-19 pandemic in Taiwan, and most confirmed patients underwent screening and evaluation in EDs [29]. To provide essential care with optimized infection control in the ED, a double triage and telemedicine protocol was developed at our hospital. This study was conducted to evaluate the benefits and feasibility of the novel protocol for reducing the risk of COVID-19 infection among health care workers who manage COVID-19 patients.

Methods

Study Setting and Design

The National Taiwan University Hospital (NTUH) is a 2700-bed teaching hospital that provides both primary and tertiary care. It was designated as a COVID-19 response hospital in Taiwan. From March to April 2020, the number of average daily visits was 211, and the number of average daily visits to the NTUH ED by patients posing a risk of COVID-19 transmission was 12.

We conducted a retrospective study using prospectively collected data of patients who visited the ED between March 1 and April 30, 2020. The study was approved by the Institutional Review Board of the NTUH (REC No. 202003043RINA), and the requirement to obtain informed consent was waived.

The Double Triage and Telemedicine Protocol

In the course of the health care response to the COVID-19 pandemic, we developed a double triage and telemedicine protocol for the ED to manage patients suspected of having COVID-19. The protocol comprised two major components:

1) streamlining and diverting patient inflow using a double triage method and 2) evaluating suspected COVID-19 patients using telemedicine (Figure 1).

Regarding the double triage method, the first triage (Triage 1) was set up outside the entrance of the ED to screen patients based on their likelihood of posing a risk of COVID-19 transmission. The double triage helped examine the patients' history of travel, occupation, contact, and cluster (TOCC) based on the guidelines issued by the Taiwan Centers for Disease Control and available epidemiological data [29]. The Triage 1 nurse used the Taiwan Triage and Acuity Scale (TTAS) to carry out a second triage (Triage 2) for patients who posed a risk of COVID-19 transmission to health care workers [30]. TTAS is a computerized triage system with a 5-level scale that classifies patients in descending order of acuity from Triage Level I (resuscitation) and Triage Level II (emergent) to Triage Level V (nonurgent). The TTAS assesses triage level according to existing chief complaints, vital signs (eg, degree of respiratory distress, hemodynamic stability, conscious level, change in body temperature, and pain severity), and mechanism of injury (for patients with trauma) to determine the triage level of a patient. The patients who triaged as TTAS level I usually had life-threatening diseases and needed immediate resuscitation. Those patients who triaged as TTAS level II usually had potential life-threatening diseases or organ dysfunction and needed management quickly. The recommended reassessment time was less than 10 minutes. For patients triaged as TTAS levels III to V, the recommended reassessment times were 30, 60, and 120 minutes, respectively. If a patient was triaged as TTAS Level I or II, they would be directed to an isolated COVID-19 resuscitation room for further management. Patients classified as TTAS levels III to V were referred to the COVID-19 emergency clinic, which is located in a structure that is separated from the main ED building. The COVID-19 resuscitation rooms and COVID-19 emergency clinics are equipped with new ED infrastructure that was constructed in response to emerging infectious diseases [31]. Patients with no risk of COVID-19 would be diverted by the Triage 1 nurse into an internal triage facility inside the ED and receive routine emergency care.

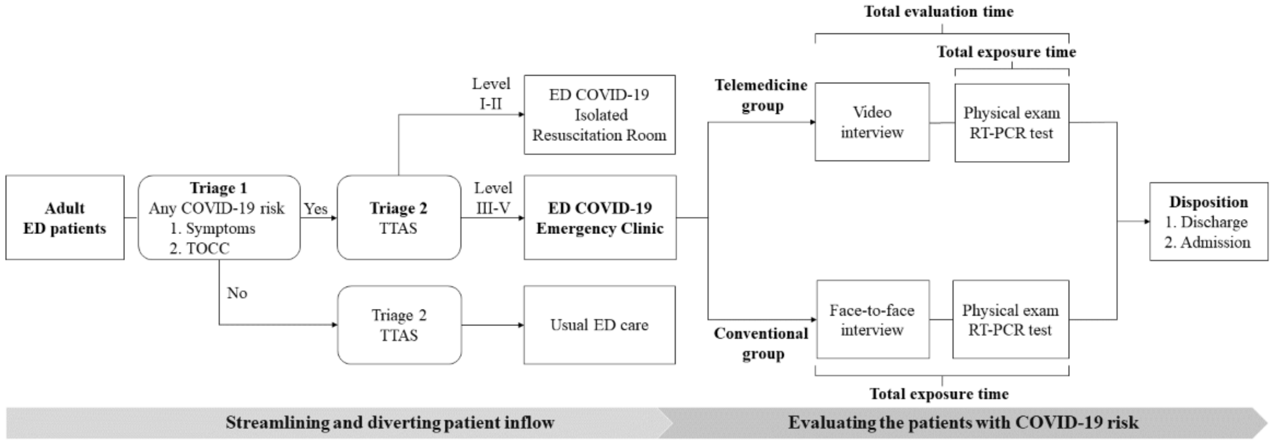
The health care workers in the COVID-19 emergency clinic were instructed to don appropriate PPE, including an N95 face mask, a waterproof gown, a non-disposable face shield, a hair cap, shoe sleeves, and two layers of gloves, when working in the clinic. A nurse was assigned to verify the appropriateness of the PPE and record the times of donning PPE, doffing PPE, and entering and exiting the clinic. In the COVID-19 emergency clinic, physicians performed face-to-face interviews with patients before performing physical examinations and collecting specimens for COVID-19 reverse transcription–polymerase chain reaction (RT-PCR) tests. In our study, patients evaluated using face-to-face interviews were assigned to the conventional group.

To improve infection control in the ED, we established a COVID-19 telemedicine team of nine attending physicians and four senior residents and provided them with training in the telemedicine interview system (TIS) on March 27, 2020. On April 1, 2020, the TIS was introduced in the COVID-19 emergency clinic (Multimedia Appendix 1) using U Meeting (CyberLink Corp) as the communication software. The TIS protects the privacy and security of patient data through end-to-end encryption and data transmission through an intranet governed by the firewall of the hospital's information system. Furthermore, the TIS is in compliance with the government regulations for telemedicine. If a patient who was diverted to the COVID-19 emergency clinic agreed to attend a video interview, the patient's attending physician on the COVID-19 telemedicine team would first use the TIS to conduct that interview. The physician would then share the patient's chest radiography findings and disease information on a screen located at the patient's side while explaining them to the patient. After the TIS interview, the physician would don appropriate PPE before entering the clinic to complete all the other necessary evaluations and obtain specimens for the COVID-19 RT-PCR test. We designated patients who underwent TIS evaluation as the telemedicine group. In both the telemedicine group and the conventional group, physicians would then decide the disposition of patients (ie, either discharging them or admitting them to quarantine wards).

The total exposure time was defined as the time that a physician remained in contact with a patient during their face-to-face interview. The total evaluation time was defined as the time the physician took to interview and assess the patient and collect specimens for laboratory testing. Both these time intervals were prospectively recorded.

Furthermore, we conducted a survey using a 10-item questionnaire in both groups upon completion of the clinical evaluation. The questionnaire was modified from a questionnaire for evaluating patient satisfaction with telemedicine [32]. It was also simplified to assess the immediate impression of the study participants because the survey was conducted in an isolated room and the response time was limited. The questionnaire helped us obtain information on the global rating of the interview, the quality of the interview, the mutual understanding between the patient and the physician, and perceptions of stress, discrimination, and safety. To assess the feasibility of telemedicine, we asked 3 additional questions to survey the participants' acceptance, perception of safety, and satisfaction with the evaluation protocol. The level of agreement was assessed using the 5-point Likert scale, ranging from 1 point (extremely agree) to 5 points (extremely disagree). The questionnaire was validated by 4 emergency physicians and the director of the Center of Quality Management at the study center. Data on the patients' admission, discharge, and clinical and personal backgrounds were obtained from the hospital's health information system.

Figure 1. Diagram of the double triage and telemedicine protocol. COVID-19: coronavirus disease; ED: emergency department; TOCC: travel, occupation, contact, and cluster history; TTAS: Taiwan Triage and Acuity Scale.



Recruitment

We included adult patients aged ≥20 years who were diverted to our COVID-19 emergency clinic during the study period. Patients who refused TIS in the telemedicine group were excluded from this study. Furthermore, to control the confounding effect of evaluation by different physicians, patients not evaluated by the COVID-19 telemedicine team were excluded from the analysis dataset.

Outcomes

The primary outcomes of interest included physicians' total exposure time and total evaluation time. The secondary outcomes were the mean scores of the questionnaire survey used to evaluate the double triage and telemedicine protocol.

Statistical Analysis

Descriptive statistics are presented as mean (SD) for continuous variables, and the intergroup differences in means were analyzed using the independent sample *t* test. The chi-square test was used to assess the associations between categorical variables, namely sex, marital status, education, TOCC history, chronic health conditions, and primary care physician level. To compare the total exposure time and total evaluation time between the two groups, we calculated their intergroup mean differences and 95% CIs. Multiple linear regression models were constructed to adjust for potential confounders and determine the effect of the telemedicine group on each of the outcomes. The identified covariates included age, sex, education, TTAS level, and primary care physician level. The goodness of fit for the multiple linear regression models was examined by computing the R^2 statistic. Regression diagnostics were used to identify problems in the models or data. The answers to each questionnaire item were analyzed using the Mann-Whitney U test and are presented as mean scores. A two-tailed *P* value ≤.05 was considered to be statistically significant. All statistical analyses were conducted using SAS version 9.4 (SAS Institute).

Results

Characteristics of the Study Participants

We initially enrolled 707 adult ED patients in the study, of whom 203 (28.7%) were excluded from the study because of high TTAS levels (I and II). Before implementing the TIS, 342 patients were interviewed, of whom 237 (69.3%) were not interviewed by the telemedicine team. After the introduction of the TIS, 162 patients were interviewed, including 1 patient who refused the TIS and 68 others who were not interviewed by the telemedicine team. We also excluded all patients who were not interviewed by the telemedicine team and those who refused the TIS. Finally, 93 patients were included in the telemedicine group for analysis, compared with 105 patients in the conventional group (Figure 2).

The clinical characteristics of the patients are presented in Table 1. Compared to the conventional group, the telemedicine group had significantly higher rates of patients with an education level of senior high school or lower (31/93, 33.3%, vs 16/105, 15.2%; *P*=.003), TTAS triage level III (76/93, 81.7%, vs 58/105, 55.2%; *P*<.001), and cardiovascular diseases (8/93, 8.6%, vs 1/105, 1.0%; *P*=.014). The telemedicine group had a lower rate of patients with a travel history than the conventional group (26/93, 28.0%, vs 56/105, 53.3%; *P*<.001). The rate of resident doctors conducting interviews with patients in the telemedicine group was higher than that in the conventional group (39/93, 41.9%, vs 17/105, 16.2%; *P*<.001).

The mean age showed no significant difference between the telemedicine group and the conventional group (mean 39.8 years, SD 17.1, vs mean 38.4 years, SD 25.1; *P*=.65). The two groups did not show any difference in terms of the rate of male patients (41/93, 44.1%, in the telemedicine group vs 44/105, 41.9%, in the conventional group; *P*=.76) or that of married patients (22/93, 23.7%, in the telemedicine group vs 21/105, 20%, in the conventional group; *P*=.53). Regarding comorbidities, the rates of patients with diabetes (8/93, 8.6%, vs 2/105, 1.9%; *P*=.048), chronic renal disease (4/93, 4.3%, vs. 0/105, 0.0%; *P*=.047), and cardiovascular diseases (8/93, 8.6%, vs 1/105, 1.0%; *P*=.014) were significantly higher in the telemedicine group than in the conventional group. Meanwhile, no significant discrepancies were observed in the other rates.

Figure 2. Flowchart of patient disposition. COVID-19: coronavirus disease; ED: emergency department; TTAS: Taiwan Triage and Acuity Scale.

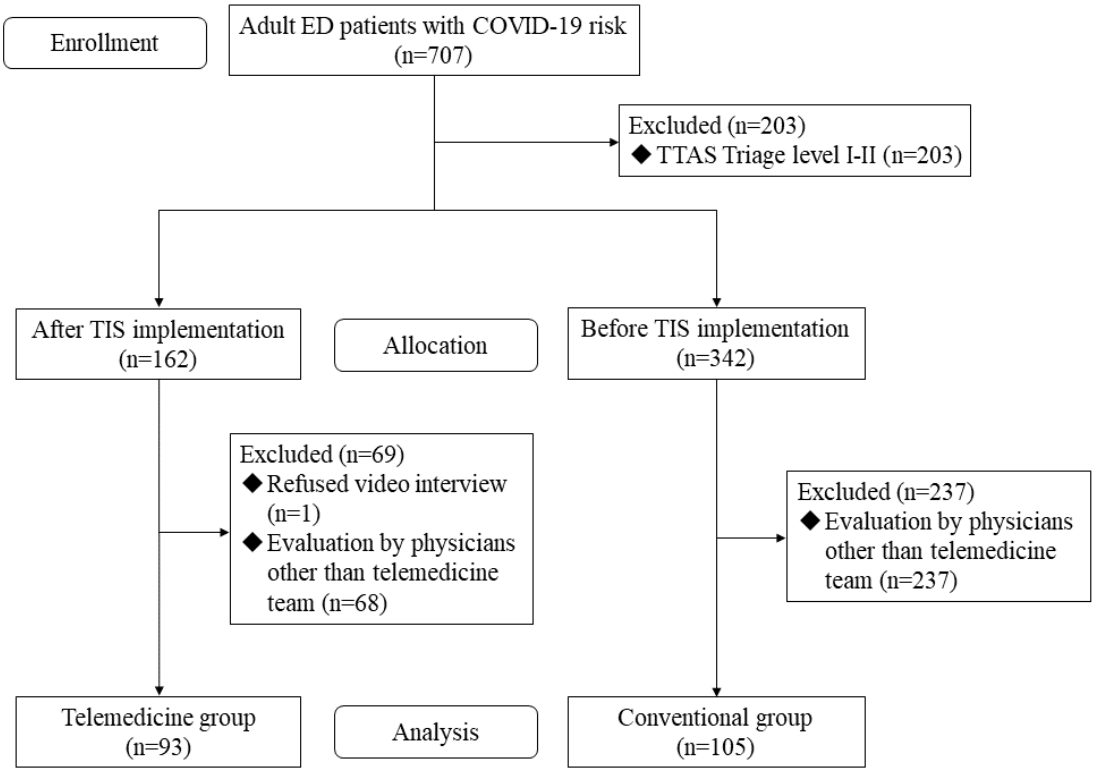


Table 1. Baseline demographic and clinical characteristics of patients in the telemedicine and conventional groups (N=198).

Characteristic	Telemedicine group (n=93)	Conventional group (n=105)	*P* value
Age (years), mean (SD)	39.8 (17.1)	38.4 (25.1)	.65
Male gender, n (%)	41 (44.1)	44 (41.9)	.76
Marital status, n (%)			**.53**
Married	22 (23.7)	21 (20.0)	
Single, divorced, or widowed	71 (76.3)	84 (80.0)	
Education, n (%)			**.003**
Senior high school or lower	31 (33.3)	16 (15.2)	
University or higher	62 (66.7)	89 (84.8)	
Epidemiologically significant TOCC[a] history, n (%)			
Travel	26 (28.0)	56 (53.3)	<.001
Occupation	36 (38.7)	41 (39.1)	.96
Contact	19 (20.4)	25 (23.8)	.57
Cluster	1 (1.1)	2 (1.9)	>.999
ED[b] triage level, n (%)			**<.001**
Triage level III	76 (81.7)	58 (55.2)	
Triage level IV or V	17 (18.3)	47 (44.8)	
Comorbidities, n (%)			
Diabetes	8 (8.6)	2 (1.9)	.05
Malignancy	4 (4.3)	1 (1.0)	.19
Chronic renal disease	4 (4.3)	0 (0.0)	.05
Chronic liver disease	2 (2.2)	1 (1.0)	.60
Cardiovascular diseases	8 (8.6)	1 (1.0)	.01
COPD[c] or asthma	3 (3.2)	5 (4.8)	.73
Cerebrovascular accident	1 (1.1)	0 (0.0)	.47
Hypertension	11 (11.8)	6 (5.7)	.13
Primary care physician level, n (%)			**<.001**
Attending physician	54 (58.1)	88 (83.8)	
Resident	39 (41.9)	17 (16.2)	
ED disposition, n (%)			**.06**
Admission	20 (21.5)	35 (33.3)	
Discharge	73 (78.5)	70 (66.7)	
72-hour ED revisit,[d] n (%)	2 (2.7)	0 (0.0)	.50

[a] TOCC: travel/occupation/contact/cluster.

[b] ED: emergency department.

[c] COPD: chronic obstructive pulmonary disease.

[d] Hospitalized patients are excluded.

Comparison of the Total Exposure Time, Total Evaluation Time, and 72-Hour Revisit Rate Between the Telemedicine Group and the Conventional Group

The total exposure time and total evaluation time were compared to estimate the benefits of the implementation of the double triage and telemedicine protocol (Figure 3). The total exposure time in the telemedicine group was significantly shorter than that in the conventional group (4.7 minutes, SD 2.4, vs. 8.9 minutes, SD 4.3; *P*<.001). In contrast, the total evaluation time in the telemedicine group was longer than that in the conventional group (12.2 minutes, SD 3.5, vs 8.9 minutes, SD 4.3; *P*<.001). To evaluate the quality of interviews, the two groups were compared in terms of the 72-hour ED revisit rate.

The analysis did not show a statistically significant difference in this rate between the groups (2/93, 2.7%, in the telemedicine group vs 0/105, 0.0%, in the conventional group; *P*=.50) (Table 1). The crude and adjusted mean difference estimates of the total exposure time and total evaluation time between the two groups are shown in Table 2. After adjusting for age, gender, triage level, educational status, and primary care physician level, we found that the total exposure time in the telemedicine group was 4.6 minutes shorter (95% CI −5.7 to −3.5, *P*<.001) than that in the conventional group. However, the total evaluation time in the telemedicine group was 2.8 minutes longer (95% CI −1.6 to −4.0; *P*<.001) than in the conventional group.

Figure 3. Box plot of the differences in the total exposure time and the total evaluation time between the telemedicine and conventional groups.

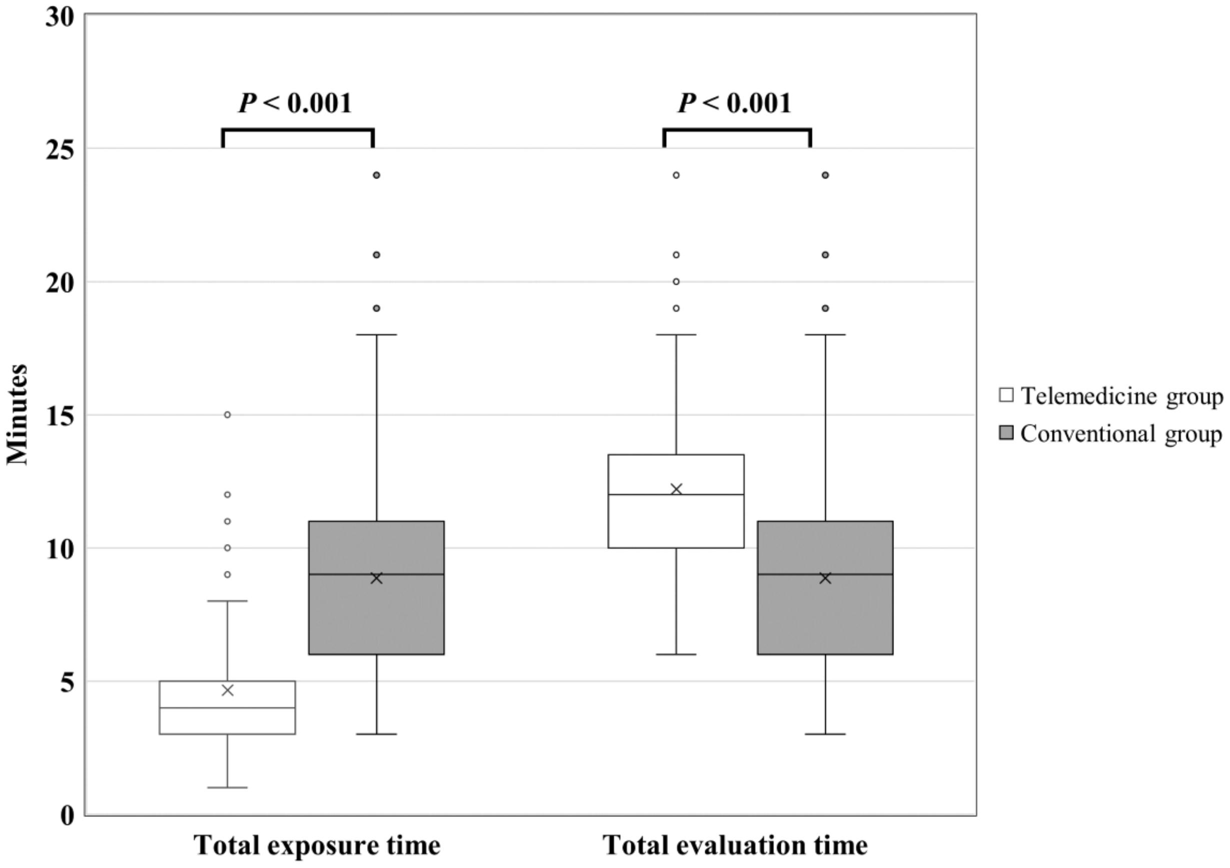


Table 2. Estimates of the crude and adjusted mean differences in the total evaluation time and total exposure time in minutes between the telemedicine and conventional groups.

Outcome	Telemedicine group (n=93), mean (SD)	Conventional group (n=105), mean (SD)	Mean difference estimate[a]		95% CI		*P* value	
			Crude	Adjusted[b]	Crude	Adjusted	Crude	Adjusted
Total exposure time	4.7 (2.4)	8.9 (4.3)	–4.2	–4.6	–5.2 to –3.2	–5.7 to –3.5	<.001	<.001
Total evaluation time	12.2 (3.5)	8.9 (4.3)	3.3	2.8	2.2 to 4.4	1.6 to 4.0	<.001	<.001

[a] The conventional group estimate was subtracted from the telemedicine group estimate.

[b] The model was adjusted for age, gender, triage level, educational status, and primary care physician level.

Comparison of the Mean Scores of the Questionnaire Survey Between the Telemedicine Group and the Conventional Group

Table 3 compares the mean scores of the questionnaire survey between the telemedicine group and the conventional group. Regarding the quality of interpreting imaging and laboratory studies (Question 4), the mean scores for the telemedicine group and the conventional group were 4.7/5 and 4.5/5, respectively (*P*=.07). The scores in relation to acceptance, safety, and satisfaction with the telemedicine protocol were all 4.7/5. The overall patient satisfaction with ED visits did not significantly differ between the two groups (mean score 4.6 vs 4.5, *P*=.33). There were no intergroup differences in the other variables (Table 3).

Table 3. Numbers of respondents by point of the 5-point Likert scale and mean scores of the telemedicine and conventional groups.

Survey question	Telemedicine group (n=82)[a]						Conventional group (n=96)[b]						*P* value[c]
	Response[d], n (%)					Mean score	Response, n (%)					Mean score	
	1	2	3	4	5		1	2	3	4	5		
1. I am satisfied with the visit.	1 (1)	0 (0)	3 (4)	20 (24)	58 (71)	4.6	0 (0)	0 (0)	3 (3)	29 (30)	60 (63)	4.5	.33
2. I had enough time to tell the doctor about what happened to me.	1 (1)	0 (0)	1 (1)	21 (26)	59 (72)	4.7	0 (0)	0 (0)	4 (4)	29 (30)	63 (66)	4.6	.35
3. The doctor understood my presentation well.	0 (0)	0 (0)	6 (7)	20 (24)	56 (68)	4.6	0 (0)	0 (0)	6 (6)	29 (30)	61 (64)	4.6	.58
4. The doctor clearly explained evaluation and X-ray results to me.	0 (0)	1 (1)	3 (4)	16 (20)	62 (76)	4.7	0 (0)	2 (2)	8 (8)	25 (26)	61 (64)	4.5	.07
5. I could hear the doctor's voice clearly.	0 (0)	0 (0)	2 (2)	23 (28)	57 (70)	4.7	0 (0)	1 (1)	5 (5)	24 (25)	66 (69)	4.6	.78
6. I had enough time to ask questions.	1 (1)	0 (0)	3 (4)	21 (26)	57 (70)	4.6	0 (0)	2 (2)	4 (4)	28 (29)	62 (65)	4.6	.48
7. I felt relaxed when I talked to the doctor.	1 (1)	0 (0)	3 (4)	19 (23)	59 (72)	4.6	0 (0)	1 (1)	7 (7)	21 (22)	67 (70)	4.6	.67
8. During the visit, I was not scared or stressed.	1 (1)	0 (0)	3 (4)	18 (22)	60 (73)	4.7	0 (0)	3 (3)	4 (4)	26 (27)	63 (66)	4.6	.27
9. I did not feel discriminated during the visit.	1 (1)	0 (0)	2 92)	17 (21)	62 (76)	4.7	0 (0)	0 (0)	4 (4)	25 (26)	67 (70)	4.7	.40
10. My privacy was well-protected.	0 (0)	1 (1)	6 (7)	19 (23)	56 (68)	4.6	0 (0)	0 (0)	7 (7)	27 (28)	62 (65)	4.6	.69
11. I think video interviews are acceptable.	0 (0)	0 (0)	2 (2)	22 (27)	58 (71)	4.7	N/A[e]	N/A	N/A	N/A	N/A	N/A	N/A
12. I felt safe during the video interview.	1 (1)	0 (0)	4 (5)	15 (18)	62 (76)	4.7	N/A	N/A	N/A	N/A	N/A	N/A	N/A
13. I am satisfied with the video interview.	1 (1)	0 (0)	2 (2)	18 (22)	61 (74)	4.7	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[a] A total of 11 non-respondents in the telemedicine group were excluded.

[b] A total of 9 non-respondents in the conventional group were excluded.

[c] Analyzed using the Mann-Whitney U test.

[d] 5-point Likert scale: 1. Strongly disagree. 2. Disagree. 3. Neutral. 4. Agree. 5. Strongly agree.

[e] Not applicable.

Discussion

Principal Results

This study shows that the double triage and telemedicine protocol in the ED could reduce physicians' time of exposure to patients who pose a risk of COVID-19 transmission without compromising patient satisfaction. The time of direct exposure to individual patients in the telemedicine group could be 39% to 64% of that in the conventional group. During the COVID-19 pandemic, our protocol could effectively protect our health care workers from contracting infections in the ED. Moreover, wearing PPE has been shown to constrict mobility and vision and cause heat stress and dehydration, especially in hot weather [33]. This protocol could therefore alleviate the workload and stress of health care workers. However, the total evaluation time in the telemedicine group was longer than that in the conventional group; this may be due to the fact that physicians spent more time communicating with and providing explanations to patients. Our protocol provided a safer and more comfortable interview environment than the conventional method.

Limitations

Our study has some limitations that must be addressed. First, it was a retrospective study using prospectively collected data. Confounding factors during this period, including policy changes, patient characteristics, medical resources, and laboratory examination, may have affected the study results. However, the patients in both the telemedicine group and the conventional group were relatively comparable. The policy in

Taiwan and the criteria of reporting remained roughly unchanged, leading to small confounding effects of time and other policy factors. Second, our patients were relatively young, and telecommunication use was quite common among them; this explains their high familiarity with and acceptance of video interviews. Further studies should therefore investigate the feasibility and benefits of telemedicine among older adults. However, in our COVID-19 emergency clinic, a nurse was assigned to assist patients in receiving video calls if needed, and the patients' families were allowed to accompany them in the clinic and could also help them use the TIS. These mechanisms facilitated the efficient functioning of the system. Third, in our study, 203 patients were excluded due to high TTAS levels. The feasibility of telemedicine for these patients was not investigated. Because these patients are usually in critical condition, face-to-face emergent management may be necessary in clinical scenarios. The possible application of telemedicine with critical patients should be further developed and examined. Fourth, the study was conducted at a single ED in Taiwan. Therefore, the generalizability of the novel model has not been confirmed. Furthermore, the cultural differences between Taiwan and other countries should be considered to ascertain the effectiveness of the proposed model.

Comparison With Prior Work

The double triage method in our protocol facilitated the use of a TIS. The epidemic of an emerging infectious disease may cause a surge of patients in EDs due to clinical symptoms or fear of the disease. This may overwhelm the health care capacity of overcrowded EDs and aggravate the risk of cross infection in hospitals [34]. The CDC has suggested mandatory rapid, safe triage and isolation of patients with symptoms of COVID-19 in EDs for infection control. On admission, all patients should be surveyed about the presence of fever, symptoms of COVID-19, or contact with suspected COVID-19 cases. Patients with symptoms of COVID-19 should be isolated for examination in a separate, well-ventilated space [7]. In Taiwan, we adapted our response to the COVID-19 outbreak based on our experience with SARS in 2003, whereby early identification of patients suspected of having COVID-19 through "triage screening" could help prevent in-hospital transmission [34]. Our ED developed a double triage method to survey suspected patients with COVID-19 and separated them from others. Moreover, this double triage method helped identify patients eligible for telemedicine interviews. For suspected patients with COVID-19 in critical condition, physicians should immediately proceed to face-to-face evaluation and provide them with resuscitation. However, telemedicine interviews could be a safe alternative to face-to-face interviews for patients in stable condition. The protocol described in our study may provide a practical and feasible strategy for other EDs to improve their infection control measures.

Telemedicine has long been used to provide medical care in remote areas and has proved to be beneficial in infection control and management. The published literature demonstrates that telemedicine can increase access to care, with high patient satisfaction, improved outcomes, and reduced costs [35]. Furthermore, the Infectious Diseases Society of America has issued a position statement on telehealth that outlines the various uses of telemedicine and telehealth and supports the appropriate use of telehealth in clinical care, research, and education [36]. Some applications for infection control have been reported in the past decades [19,23], such as telemonitoring of asymptomatic individuals identified as case contacts during the Ebola virus disease outbreak in Africa in 2014. Other applications include caring for symptomatic cases that require isolation (Taiwan during the SARS epidemic in 2003 and the H1N1 influenza pandemic in 2009), tele-expert consultation, and tending patients without access to health care facilities (the MERS epidemic in Korea in 2015). In the current COVID-19 pandemic, many countries are accelerating their transformation to the virtual care interface [37]. Telemedicine, particularly video consultations, has been promoted and scaled up to reduce the risk of transmission. However, almost all existing evidence pertains to highly selected samples of outpatients with chronic, stable conditions. It is largely irrelevant to the current escalating situation that involves patients with an acute and potentially serious illness. Our study showed that telemedicine is feasible and can benefit both infection control and the provision of quality care for suggested COVID-19 patients. Our protocol involves the novel synchronous application of telemedicine for infection control.

The use of video interviews in our TIS had certain advantages over telephonic or conventional methods. Through face-to-face communication, we can directly confirm the identity of patients in routine practice and examine their general appearance and respiratory status. Moreover, the telemedicine system has a share function that can help physicians to simultaneously send high-quality images, laboratory data, and health information to their patients. Notably, our study results showed that the telemedicine group had a more favorable impression of the quality of images and interpretation of laboratory studies than the conventional group, although no statistical significance was observed.

Certain concerns exist about the feasibility and acceptance of telemedicine, especially in the context of the pandemic [24,27]. In this study, we evaluated patients' perception and acceptance of telemedicine, and the results indicated that both groups showed similar overall satisfaction with the quality of interviews. With regard to the perception of stress, discrimination, and privacy, there were no significant differences between the groups. In fact, real-time video interviews may reduce patients' anxiety and physicians' stress.

Conclusions

The implementation of a double triage and telemedicine protocol during the COVID-19 pandemic has high potential to improve infection control. Our study preliminarily validated a promising model in the ED to minimize physicians' direct contact with non-critical suspected COVID-19 patients during evaluation. During the pandemic, this model could help protect critical medical personnel in the health care system from unnecessary exposure and further prevent overwhelming of the health care system.

Authors' Contributions

CHL, WPT, JLW, JT, and MTC conceived and designed the studies. CHL, HNO, HYL, YYC, CHW, and JWC contributed to the acquisition of the data. CHL and WPT contributed to the analysis of the data. The manuscript was drafted by CHL, WPT, CCC, CHH, and SCC. CCC, CHH, and SCC supervised the trial and data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Photographs of the physician's site (A) and the patient's site (B) in the COVID-19 emergency clinic.
[PNG File , 972 KB - jmir_v22i6e20586_app1.png]

References

1. Summary of probable SARS cases with onset of illness from 1 November to 31 July 2003. World Health Organization. 2004 Apr 21. URL: https://www.who.int/csr/sars/country/table2004_04_21/en/ [accessed 2020-06-19]
2. Chan M. Influenza A(H1N1): lessons learned and preparedness. Chan M. 2009 Jul 02. URL: https://www.who.int/dg/speeches/2009/influenza_h1n1_lessons_20090702/en/ [accessed 2020-06-19]
3. Middle East respiratory syndrome coronavirus (MERS-CoV): MERS Monthly Summary, November 2019. World Health Organization. 2019 Nov. URL: http://www.who.int/emergencies/mers-cov/en/ [accessed 2020-06-19]
4. Oleribe OO, Salako BL, Ka MM, Akpalu A, McConnochie M, Foster M, et al. Ebola virus disease epidemic in West Africa: lessons learned and issues arising from West African countries. Clin Med (Lond) 2015 Feb;15(1):54-57 [FREE Full text] [doi: 10.7861/clinmedicine.15-1-54] [Medline: 25650199]
5. Novel Coronavirus – China. World Health Organization. 2020 Jan 12. URL: http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ [accessed 2020-06-19]
6. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020 May;581(7809):465-469. [doi: 10.1038/s41586-020-2196-x] [Medline: 32235945]
7. Interim Infection Prevention and Control Recommendations for Patients with Suspected or Confirmed Coronavirus Disease 2019 (COVID-19) in Healthcare Settings. US Centers for Disease Control and Prevention. 2020 May 18. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control-recommendations.html [accessed 2020-06-19]
8. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:online ahead of print. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
9. Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy. JAMA 2020 Mar 17:online ahead of print. [doi: 10.1001/jama.2020.4344] [Medline: 32181795]
10. CDC COVID-19 Response Team. Characteristics of Health Care Personnel with COVID-19 - United States, February 12-April 9, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):477-481 [FREE Full text] [doi: 10.15585/mmwr.mm6915e6] [Medline: 32298247]
11. Coronavirus disease 2019 (COVID-19) Situation Report–82. World Health Organization. 2020 Apr 11. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200411-sitrep-82-covid-19.pdf?sfvrsn=74a5d15_2 [accessed 2020-06-19]
12. Wang J, Zhou M, Liu F. Reasons for healthcare workers becoming infected with novel coronavirus disease 2019 (COVID-19) in China. J Hosp Infect 2020 May;105(1):100-101 [FREE Full text] [doi: 10.1016/j.jhin.2020.03.002] [Medline: 32147406]
13. Zundel KM. Telemedicine: history, applications, and impact on librarianship. Bull Med Libr Assoc 1996 Jan;84(1):71-79 [FREE Full text] [Medline: 8938332]
14. Heinzelmann PJ, Lugn NE, Kvedar JC. Telemedicine in the future. J Telemed Telecare 2005;11(8):384-390. [doi: 10.1177/1357633X0501100802] [Medline: 16356311]
15. Taylor P. A survey of research in telemedicine. 2: Telemedicine services. J Telemed Telecare 1998;4(2):63-71. [doi: 10.1258/1357633981931948] [Medline: 9744160]
16. Taylor P. A survey of research in telemedicine. 1: Telemedicine systems. J Telemed Telecare 1998;4(1):1-17. [doi: 10.1258/1357633981931227] [Medline: 9640705]
17. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 03;320(7248):1517-1520 [FREE Full text] [doi: 10.1136/bmj.320.7248.1517] [Medline: 10834899]
18. Lurie N, Carr BG. The Role of Telehealth in the Medical Response to Disasters. JAMA Intern Med 2018 Jun 01;178(6):745-746. [doi: 10.1001/jamainternmed.2018.1314] [Medline: 29710200]
19. Ohannessian R. Telemedicine: Potential applications in epidemic situations. Eur Res Telemed Rech Eur en Télémédecine 2015 Sep;4(3):95-98 [FREE Full text] [doi: 10.1016/j.eurtel.2015.08.002]
20. Trethewey SP, Beck KJ, Symonds RF. Video consultations in UK primary care in response to the COVID-19 pandemic. Br J Gen Pract 2020 May;70(694):228-229. [doi: 10.3399/bjgp20X709505] [Medline: 32354807]

21. John O. Video consultations for triage of patients with covid-19. BMJ 2020 Apr 23;369:m1583. [doi: 10.1136/bmj.m1583] [Medline: 32327408]
22. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
23. Keshvardoost S, Bahaadinbeigy K, Fatehi F. Role of Telehealth in the Management of COVID-19: Lessons Learned from Previous SARS, MERS, and Ebola Outbreaks. Telemed J E Health 2020 Apr 23:online ahead of print. [doi: 10.1089/tmj.2020.0105] [Medline: 32329659]
24. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
25. Elkbuli A, Ehrlich H, McKenney M. The effective use of telemedicine to save lives and maintain structure in a healthcare system: Current response to COVID-19. Am J Emerg Med 2020 Apr 07:online ahead of print [FREE Full text] [doi: 10.1016/j.ajem.2020.04.003] [Medline: 32303410]
26. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
27. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
28. Yeh M, Chang H. National Health Insurance In Taiwan. Health Aff (Millwood) 2015 Jun;34(6):1067. [doi: 10.1377/hlthaff.2015.0447] [Medline: 26056216]
29. Coronavirus disease 2019 (COVID-19). Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/en/Disease/SubIndex/ [accessed 2020-05-01]
30. Ng C, Yen Z, Tsai JC, Chen LC, Lin SJ, Sang YY, TTAS national working group. Validation of the Taiwan triage and acuity scale: a new computerised five-level triage system. Emerg Med J 2011 Dec;28(12):1026-1031. [doi: 10.1136/emj.2010.094185] [Medline: 21076055]
31. Lien W, Wu J, Tseng W, Chow-In Ko P, Chen S, Tsai M, et al. Fight COVID-19 Beyond the Borders: Emergency Department Patient Diversion in Taiwan. Ann Emerg Med 2020 Jun;75(6):785-787 [FREE Full text] [doi: 10.1016/j.annemergmed.2020.04.003] [Medline: 32334881]
32. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the Telemedicine Satisfaction Questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
33. Loibner M, Hagauer S, Schwantzer G, Berghold A, Zatloukal K. Limiting factors for wearing personal protective equipment (PPE) in a health care environment evaluated in a randomised study. PLoS One 2019;14(1):e0210775 [FREE Full text] [doi: 10.1371/journal.pone.0210775] [Medline: 30668567]
34. McDonald LC, Simor AE, Su I, Maloney S, Ofner M, Chen K, et al. SARS in healthcare facilities, Toronto and Taiwan. Emerg Infect Dis 2004 May;10(5):777-781 [FREE Full text] [doi: 10.3201/eid1005.030791] [Medline: 15200808]
35. Norris E, Dunsmuir S, Duke-Williams O, Stamatakis E, Shelton N. Protocol for the 'Virtual Traveller' cluster-randomised controlled trial: a behaviour change intervention to increase physical activity in primary-school Maths and English lessons. BMJ Open 2016 Jun 27;6(6):e011982 [FREE Full text] [doi: 10.1136/bmjopen-2016-011982] [Medline: 27354084]
36. Young JD, Abdel-Massih R, Herchline T, McCurdy L, Moyer KJ, Scott JD, et al. Infectious Diseases Society of America Position Statement on Telehealth and Telemedicine as Applied to the Practice of Infectious Diseases. Clin Infect Dis 2019 Apr 24;68(9):1437-1443. [doi: 10.1093/cid/ciy907] [Medline: 30851042]
37. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]

Abbreviations

CDC: US Centers for Disease Control and Prevention
COVID-19: coronavirus disease
ED: emergency department
MERS: Middle Eastern respiratory syndrome
NTUH: National Taiwan University Hospital
PPE: personal protective equipment
RT-PCR: reverse transcription–polymerase chain reaction
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
TIS: telemedicine interview system
TOCC: travel, occupation, contact, and cluster
TTAS: Taiwan Triage and Acuity Scale

Edited by G Eysenbach; submitted 22.05.20; peer-reviewed by H Akram; comments to author 12.06.20; revised version received 15.06.20; accepted 15.06.20; published 23.06.20.

Please cite as:
Lin CH, Tseng WP, Wu JL, Tay J, Cheng MT, Ong HN, Lin HY, Chen YY, Wu CH, Chen JW, Chen SY, Chan CC, Huang CH, Chen SC
A Double Triage and Telemedicine Protocol to Optimize Infection Control in an Emergency Department in Taiwan During the COVID-19 Pandemic: Retrospective Feasibility Study
J Med Internet Res 2020;22(6):e20586
URL: http://www.jmir.org/2020/6/e20586/
doi:10.2196/20586
PMID:32544072

Original Paper

Virtual Management of Patients With Cancer During the COVID-19 Pandemic: Web-Based Questionnaire Study

Emad Tashkandi[1,2*], MBBS; Ahmed Zeeneldin[1,3*], MD, MSc; Amal AlAbdulwahab[4*], MBBS; Omima Elemam[1,5*], MD; Shereef Elsamany[1,3*], MD; Wasil Jastaniah[2,6*], MBBS; Shaker Abdullah[6*], MBBS; Mohammad Alfayez[1,2*], MBBS; Abdul Rahman Jazieh[7*], MD; Humaid O Al-Shamsi[8,9,10], MD

[1]Department of Medical Oncology, Oncology Center, King Abdullah Medical City, Makkah, Saudi Arabia

[2]College of Medicine, Umm AlQura University, Makkah, Saudi Arabia

[3]National Cancer Institute, Cairo University, Cairo, Egypt

[4]Oncology Center, King Abdullah Medical City, Makkah, Saudi Arabia

[5]Department of Medical Oncology, Oncology Centre, Mansoura University, Mansoura University, Egypt

[6]Princess Noorah Oncology Center, King Abdul-Aziz Medical City, Ministry of National Guard – Health Affairs, Jeddah, Saudi Arabia

[7]King Saudi Bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

[8]Department of Oncology, Alzahra Hospital, Dubai, United Arab Emirates

[9]Emirates Oncology Society, Dubai, United Arab Emirates

[10]Department of Medicine, University of Sharjah, Sharjah, United Arab Emirates

*these authors contributed equally

Corresponding Author:
Emad Tashkandi, MBBS
Department of Medical Oncology
Oncology Center
King Abdullah Medical City
PO box 715
Makkah, 21955
Saudi Arabia
Phone: 966 555290061
Email: Tashkandi.e@kamc.med.sa

Abstract

Background: During the coronavirus disease (COVID-19) pandemic, patients with cancer in rural settings and distant geographical areas will be affected the most by curfews. Virtual management (telemedicine) has been shown to reduce health costs and improve access to care.

Objective: The aim of this survey is to understand oncologists' awareness of and views on virtual management, challenges, and preferences, as well as their priorities regarding the prescribing of anticancer treatments during the COVID-19 pandemic.

Methods: We created a self-administrated electronic survey about the virtual management of patients with cancer during the COVID-19 pandemic. We evaluated its clinical sensibility and pilot tested the instrument. We surveyed practicing oncologists in Gulf and Arab countries using snowball sampling via emails and social media networks. Reminders were sent 1 and 2 weeks later using SurveyMonkey.

Results: We received 222 responses from validated oncologists from April 2-22, 2020. An awareness of virtual clinics, virtual multidisciplinary teams, and virtual prescriptions was reported by 182 (82%), 175 (79%), and 166 (75%) respondents, respectively. Reported challenges associated with virtual management were the lack of physical exam (n=134, 60%), patients' awareness and access (n=131, 59%), the lack of physical attendance of patients (n=93, 42%), information technology (IT) support (n=82, 37%), and the safety of virtual management (n=78, 35%). Overall, 111 (50%) and 107 (48%) oncologists did not prefer the virtual prescription of chemotherapy and novel immunotherapy, respectively. However, 188 (85%), 165 (74%), and 127 (57%) oncologists preferred the virtual prescription of hormonal therapy, bone modifying agents, and targeted therapy, respectively. In total, 184 (83%), 183 (83%), and 176 (80%) oncologists preferred to continue neoadjuvant, adjuvant, and perioperative treatments, respectively. Overall, 118 (53%) respondents preferred to continue first-line palliative treatment, in contrast to 68 (30%) and 47 (21%) respondents indicating a preference to interrupt second- and third-line palliative treatment, respectively. For administration of virtual prescriptions, all respondents preferred the oral route and 118 (53%) preferred the subcutaneous route. In contrast, 193 (87%) did not prefer the intravenous route for virtual prescriptions. Overall, 102 (46%) oncologists responded that they would “definitely” prefer to manage patients with cancer virtually.

Conclusions: Oncologists have a high level of awareness of virtual management. Although their survey responses indicated that second- and third-line palliative treatments should be interrupted, they stated that neoadjuvant, adjuvant, perioperative, and first-line palliative treatments should continue. Our results confirm that oncologists’ views on the priority of anticancer treatments are consistent with the evolving literature during the COVID-19 pandemic. Challenges to virtual management should be addressed to improve the care of patients with cancer.

(J Med Internet Res 2020;22(6):e19691) doi:10.2196/19691

KEYWORDS

teleoncology; telemedicine; eHealth; cancer; COVID-19; public health

Introduction

In December 2019, a cluster of patients with severe pneumonia were identified in Wuhan, China, and a novel coronavirus disease (COVID-19) was discovered [1]. This disease can range from asymptomatic infection to severe respiratory distress syndrome and death. The World Health Organization has declared the COVID-19 outbreak a pandemic. As of April 25, 2020, a total of 2,719,897 cases and 187,705 confirmed deaths have been reported across more than 200 countries [2].

This highly contagious virus is characterized by rapid human-to-human transmission [3], and risk factors for mortality include older age and comorbidities [4]. Patients with cancer are susceptible to COVID-19 infections because of the immunosuppressive effect of cancer treatments like chemotherapy or surgery [5], and hence have a poorer prognosis.

Teleoncology is the application of telemedicine to oncology. It has the potential to enhance access to and improve the quality of clinical cancer care [6]. Patients from rural and distant geographical areas will be most affected by curfews. Virtual management (telemedicine) has been shown to reduce health costs and improve access to care. There are examples of successful technology applications for survivorship care, palliative care, symptoms management, and supervision of satellite anticancer infusion suites [7-9]. An updated conceptual framework of telemedicine during the COVID-19 pandemic has been defined and could be applied at larger scale to improve national public health responses [10]. Further reduction of patients’ exposure to infection could be achieved by replacing certain clinic visits with virtual clinics (via videoconference or telephone) to minimize hospital visits. This allows oncologists to defer routine follow-ups, assess patients who can continue certain anticancer treatments (such as chemotherapy, immunotherapy, targeted therapy, or hormonal therapy), and continue cancer care virtually.

Multidisciplinary tumor boards ensure the selection of high priority curative cases and improve outcomes without delays or interruptions of cancer care. This could be continued virtually, depending on availability and the capacity of the health care system.

Virtual prescription and delivery of drugs is an alternative way to manage patients with cancer, especially when delivering drugs to their home via courier services or to health facilities near their home to avoid interruption of treatments, provided that this service is logistically feasible and available.

Oncologists need to weigh the risks and benefits of anticancer treatments during the pandemic. Caring for patients with cancer during this period is challenging. Jeopardizing safety by exposing patients to infection when they leave their home to visit oncology clinics and receive anticancer treatments may lead to greater risks of potential adverse events.

There is a limited number of studies to guide oncologists on how to manage patients with cancer during a pandemic. In this survey, we aim to report the views of oncologists on virtual management (awareness, challenges, and preferences) and their priorities when prescribing anticancer treatments during the COVID-19 pandemic. This could help oncologists conduct future controlled studies or trials, and guide health systems on areas of improvement in supportive infrastructure.

Methods

Study Design

This study presents the findings of a web-based questionnaire submitted to licensed oncologists in the Gulf and Arab regions.

Study Population

We included study subjects who met the following criteria: licensed practicing oncologists in the Gulf or Arab regions who are treating adult or pediatric patients, and involved in the care of patients with cancer using anticancer treatments (eg, chemotherapy, novel immunotherapy, targeted therapy, hormonal therapy, and bone modifying agents). Exclusion criteria were nononcologists and trainees.

Study Procedures

We used a nonprobability snowball sampling [11] design. To identify our target population, we contacted oncologists who are members of established national oncology associations and societies in the region to distribute and participate in the survey. If this was not applicable, we contacted 1 to 2 regional oncologists per area to distribute and participate in the survey. We used email and WhatsApp, the most popular social media network in the region, to reach oncologists.

Development of the Instrument

We generated our survey instrument using rigorous survey development and testing methods [12]. Items were selected based on a literature review, as well as email and telephone correspondence. In total, 5 experts in the field of oncology and hematology from 3 countries extensively discussed the topic and reviewed items until no further questions were missed. Items were nominated and then ranked by expert oncologists to reach a consensus on the selected items. Further review was done by methodology and content experts to eliminate redundant items using binary responses (exclude and include). We aimed to have a survey that is simple, succinct, and easy to understand. During the construction of the survey, we grouped the items into domains we wanted to explore and then refined the questions [13]. The self-administered survey (Multimedia Appendix 1) consisted of 20 items that focused on 4 domains: characteristics of oncologists; COVID-19 pandemic measures; virtual management and oncologists' views on virtual management; and the priority of prescribing anticancer treatments.

The structured response formats used in this survey included binary (yes/no), nominal, and ordinal responses. Other options were also allowed, including "Undetermined," "Other," and any other comments with free text to capture unanticipated responses. Respondents received electronic links accompanied with concise instructions, a cover letter is stating the background, the objectives of the survey, the target population, and a request to participate voluntarily (that stated their answers will be kept anonymously using SurveyMonkey).

Testing of the Instrument

During pretesting and pilot testing, questions were reviewed by 3 colleagues specializing in oncology to check the consistency and appropriateness of the questions designed by investigators [14,15], and were then reviewed by nonexpert colleagues to assess the dynamics, flow, and accessibility. In total, 5 oncology members carried out pilot testing of the instrument with minor modifications. We also conducted a clinical sensibility assessment to evaluate the comprehensiveness, clarity, and face validity of our instrument on a scale of 1 to 5. For this assessment, we invited 5 colleagues with methodologic and oncology expertise. The results of the clinical sensibility assessment, which used the mean scores indicated on a 5-point scale, suggested that the instrument had face validity (4.4), content validity (4.2), clarity (4.6), and discriminability (4.3).

Administration of the Instrument

After the approval of the King Abdullah Medical City Institutional Research Board, we sent the questionnaires electronically to licensed oncologists in the region who treat adult or pediatric patients. Oncologist types included medical oncologists, malignant hematologists, pediatric oncologists, clinical oncologists, and hemato-oncologists.

Study Duration and Timeline

On April 2, 2020, we sent participants an embedded link to the web-based survey on SurveyMonkey (along with an electronic cover letter with instructions to complete the survey) via emails, text messages, and social media such as Facebook, Twitter, and WhatsApp. Primary investigators contacted the oncology members of national associations and societies in the region to participate and create a broad distribution network. Regional oncologists distributed the survey link to their regional members and network; there were no incentives provided. We sent reminders 1 and 2 weeks later, and we closed the survey on April 22.

Statistical Analysis

Descriptive statistics were used to summarize data, and synthesize and report the views of oncologists. Description of the data also included proportions, frequencies, means, and standard deviation for continuous variables when appropriate.

Results

We received 222 completed surveys from 10 different countries in the region (Table 1). Overall, 71% (n=157) of respondents were males. Respondents have been in oncology practice for a median of 10 years. The top respondent specialty was medical oncology (n=97, 44%). Saudi Arabia is the country of current practice for 47% (n=105) of respondents. The remaining respondents practice in Arab countries. In total, 74% (n=163) practice in the public health sector and 97% (n=215) practice in urban locations.

The 222 respondents were asked whether there are any diagnosed COVID-19 cases in their country, city, hospital, and department (Table 2). In total, 97% (n=215), 97% (n=215), and 77% (n=172) of respondents indicated that there were cases in the country, city, and hospital in which they practice, respectively. In total, 18% (n=41) of respondents reported COVID-19 cases among their own patients.

Overall, out of 222 respondents, 91% (n=210) regularly attend multidisciplinary tumor boards with a monthly multidisciplinary tumor board number of 4 or more reported by 56% (n=125). In total, 82% (n=182), 79% (n=175), and 75% (n=166) of respondents were aware of virtual clinics, virtual multidisciplinary tumor boards, and virtual prescription, respectively (Table 3). Additionally, 59% (n=131), 64% (n=142), and 64% (n=143) of respondents have been personally involved in a virtual clinic, virtual multidisciplinary tumor board, and virtual prescription and delivery of drugs, respectively. Challenges faced by respondents regarding virtual management were the lack of physical examination (n=134, 60%), patients' awareness and access (n=131, 59%), the lack of physical attendance of patients (n=93, 42%), information technology (IT) support (n=82, 37%), and safety (n=78, 35%). Overall, 5% (n=10) had other comments such as the lack of a direct doctor-patient encounter, medicolegal aspects, psychological support, and privacy (Figure 1).

Table 1. Characteristics of respondents (N=222).

Characteristics	Respondents
Gender, n (%)	
Male	157 (71)
Female	65 (29)
Years in oncology practice, median	10
Specialty, n (%)	
Medical oncologist	97 (44)
Hematologist	31 (14)
Oncologist and hematologist	29 (13)
Clinical oncologist	28 (13)
Pediatric oncologist	27 (12)
Other	10 (5)
Practicing country, n (%)	
Saudi Arabia	105 (47)
United Arab Emirates	38 (17)
Egypt	18 (8)
Tunisia	13 (6)
Kuwait	10 (5)
Lebanon	8 (4)
Bahrain	7 (3)
Oman	5 (2)
Jordan	4 (2)
Other	12 (5)
Practice setting, n (%)	
Public health care	163 (73)
Private health care	33 (15)
Both public and private	26 (12)
Practice location, n (%)	
Urban	215 (97)
Rural	7 (3)

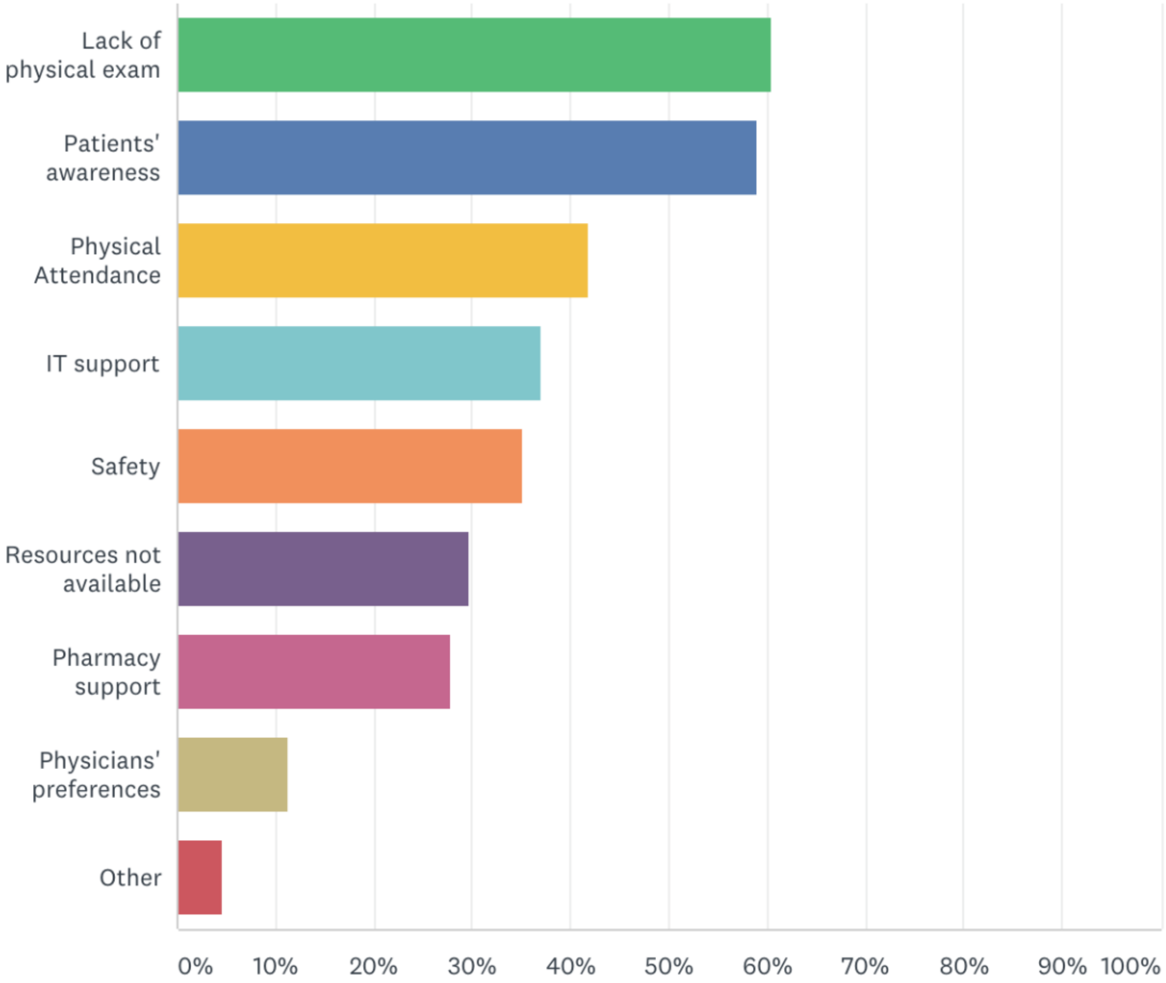
Table 2. Respondents' responses to questions about confirmed coronavirus disease cases.

Questions	Responses		
	Yes, n (%)	No, n (%)	Unknown, n (%)
Are there coronavirus disease cases in the country in which you are practicing?	215 (97)	6 (2.5)	1 (0.5)
Are there coronavirus disease cases in the city in which you are practicing?	215 (97)	6 (2.5)	1 (0.5)
Are there coronavirus disease cases in the hospital in which you are practicing?	172 (77)	41 (18)	9 (4)
Are there coronavirus disease cases in your department?	54 (24)	152 (68)	16 (7)
Are there coronavirus disease cases among your own patients?	41 (18)	165 (74)	16 (7)

Table 3. Respondents' reported awareness about virtual management.

Virtual management type	Awareness	Personally involved
Virtual clinic	182 (82)	131 (59)
Virtual tumor board	175 (79)	142 (64)
Virtual prescription and delivery of drugs	166 (75)	143 (64)

Figure 1. Respondents' reported challenges regarding virtual management. Respondents were requested to select more than one response, if applicable. IT: information technology.



When asked about the priority of prescribing anticancer treatments during the COVID-19 pandemic, 50% (n=111) and 48% (n=107) of the 222 surveyed oncologists indicated they did not prefer the virtual prescription of chemotherapy and novel immunotherapy, respectively (Figure 2). However, 85% (n=188), 74% (n=165), and 57% (n=127) of oncologists preferred the virtual prescription of hormonal therapy, bone modifying agents, and targeted therapy, respectively (Table 4).

When prescribing treatments virtually, all 222 respondents preferred treatments that are administered by the oral route and 53% (n=118) preferred the subcutaneous route. In contrast, 87% (n=193) of oncologists did not prefer the intravenous route for virtual prescriptions (Figure 3).

Of 222 respondents, more than 80% of oncologists preferred to continue neoadjuvant (n=184, 83%), adjuvant (n=183, 83%), and perioperative (n=176, 80%) treatments (Table 5). In addition, 53% (n=118) preferred to continue first-line palliative treatment. In contrast, 30% (n=68) and 21% (n=47) preferred to interrupt second- and third-line palliative treatment, respectively (Figure 4).

When the 222 oncologists were asked if they prefer to manage cases virtually, 46% (n=102) responded "Definitely," 30% (n=67) responded "Probably," 10% (n=22) answered "Neutral," 11% (n=25) said "Probably not," and 3% (n=6) said "Definitely not" (Figure 5). In total, 40% (n=87) of respondents reported that patients were satisfied with virtual management, while 18% (n=40) indicated patients were not satisfied, and 43% (n=95) answered "I don't know." Overall, 36% (n=80) of respondents indicated they are likely to continue virtual management after the pandemic, while 51% (n=112) said they will not, and 14% (n=30) answered "I don't know."

Figure 2. Respondents’ reported anti-cancer treatments that can be prescribed virtually.

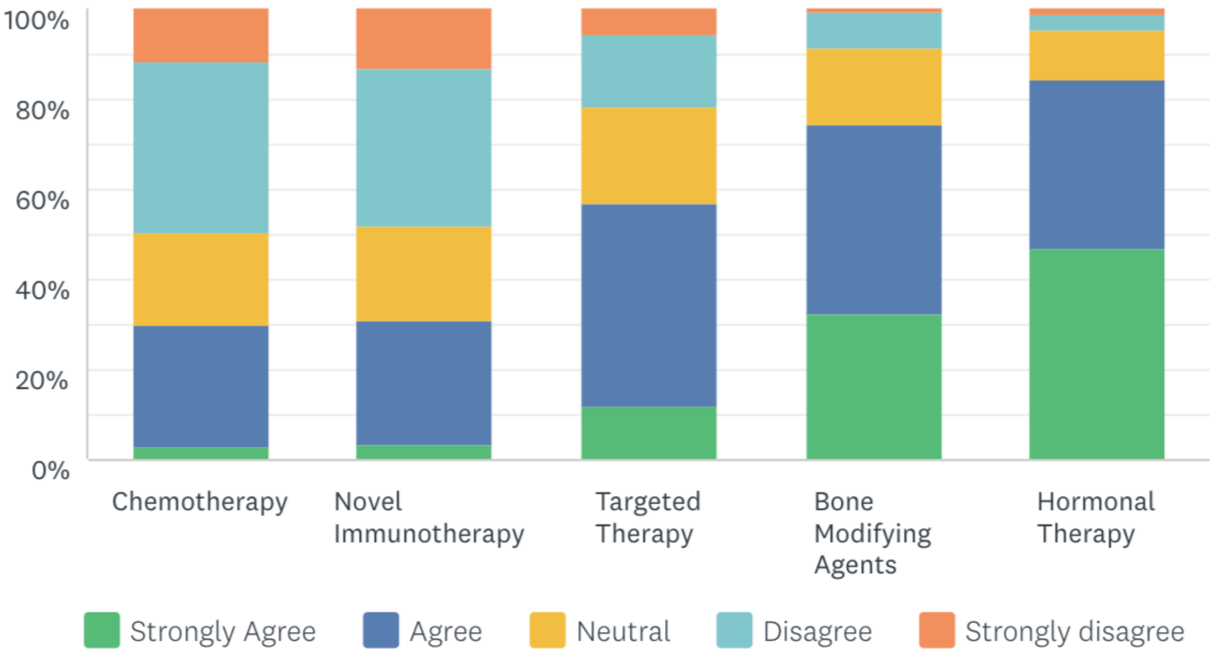


Table 4. Respondents’ reported anticancer treatments that can be prescribed virtually.

Anticancer treatments	Strongly Agree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Strongly Disagree, n (%)
Chemotherapy	7 (3)	60 (27)	45 (20)	84 (38)	26 (12)
Novel immunotherapy	8 (4)	61 (27)	46 (21)	78 (35)	29 (13)
Targeted therapy	27 (12)	100 (45)	47 (21)	36 (16)	12 (5)
Hormonal therapy	104 (47)	84 (38)	24 (110)	8 (4)	2 (10)
Bone modifying agents	72 (32)	93 (42)	38 (17)	18 (8)	1 (0.4)

Figure 3. Respondents’ reported anti-cancer treatments that can be prescribed virtually, by route. IV: intravenous. SC: subcutaneous.

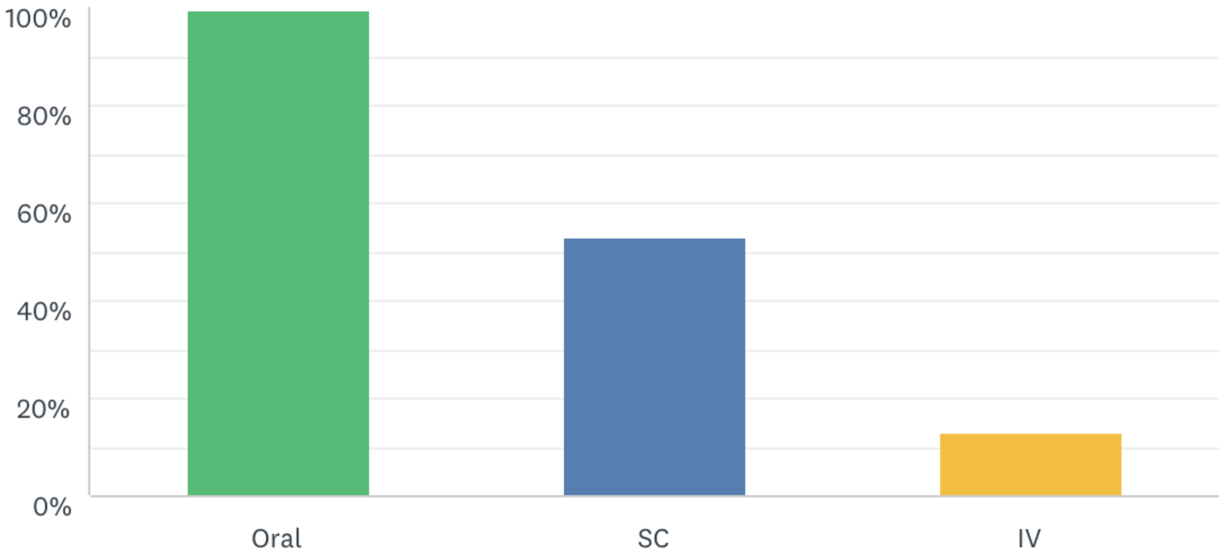


Table 5. Respondents' reported anticancer treatments that should not be interrupted.

Treatments	Strongly Agree, n (%)	Agree, n (%)	Neutral, n (%)	Disagree, n (%)	Strongly Disagree, n (%)
Neoadjuvant	114 (51)	70 (32)	24 (11)	10 (5)	4 (2)
Adjuvant	77 (35)	106 (48)	29 (13)	8 (4)	2 (1)
Perioperative	66 (30)	110 (50)	34 (15)	10 (5)	2 (1)
First-line palliative	25 (11)	93 (42)	75 (34)	27 (12)	2 (1)
Second-line palliative	12 (5)	56 (25)	87 (39)	56 (25)	11 (5)
Third-line palliative	10 (5)	37 (17)	64 (29)	73(33)	38 (17)

Figure 4. Respondents' responses regarding which anti-cancer treatments should not be interrupted.

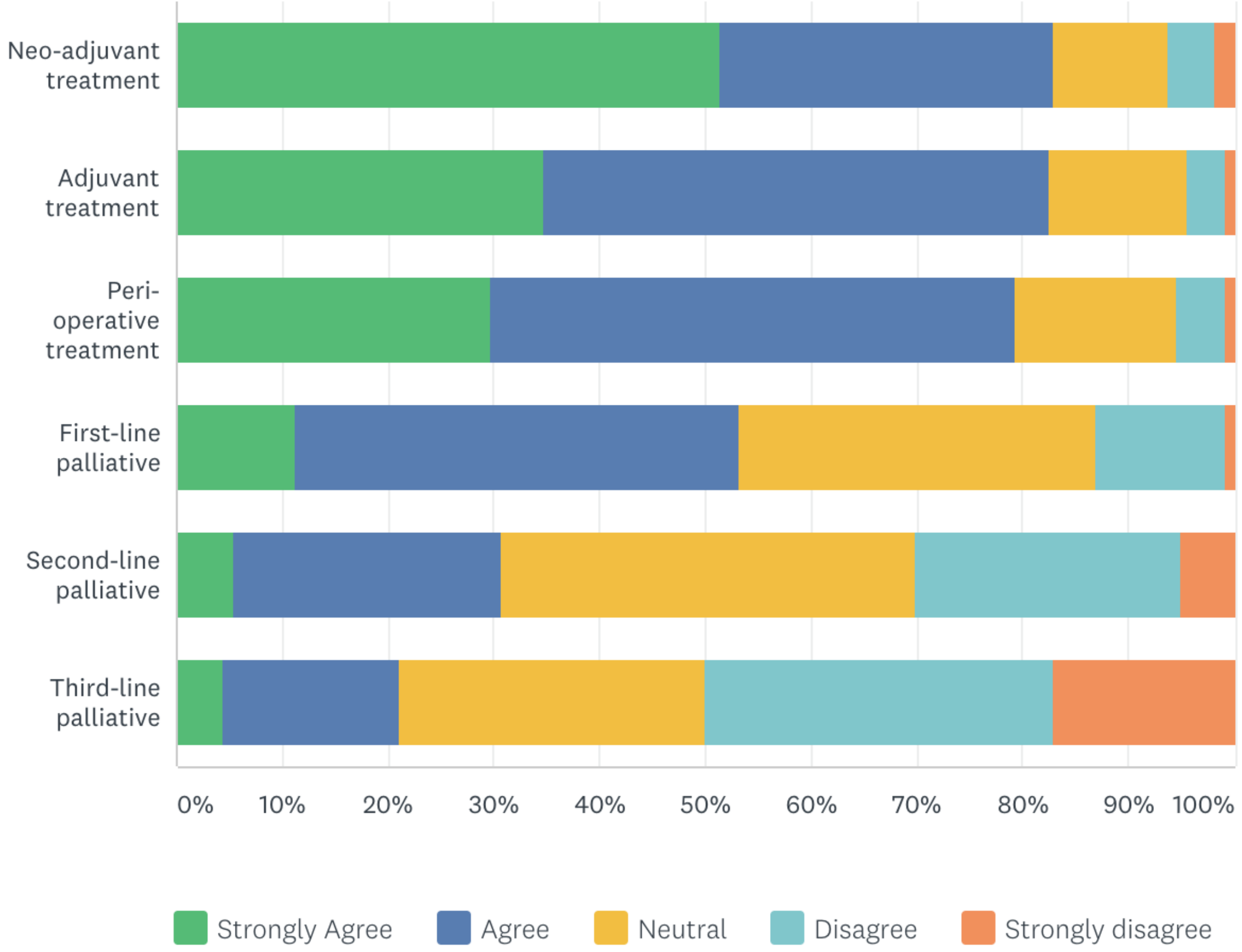
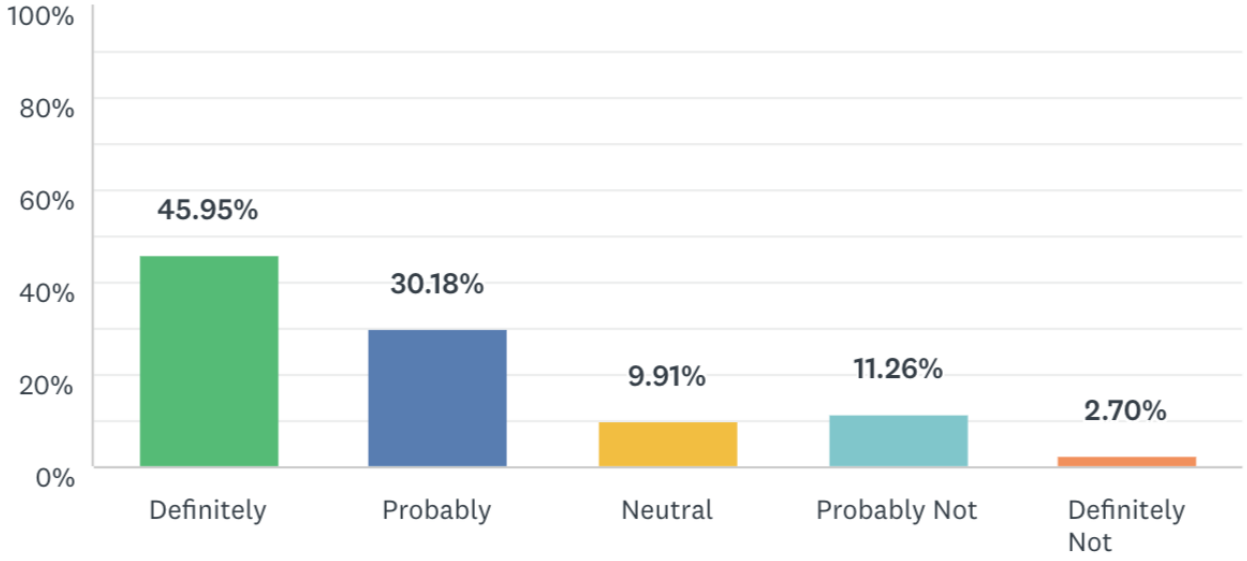


Figure 5. Respondents' responses when asked if they prefer to manage cancer patients virtually.



Discussion

Overview

During the COVID-19 curfew, oncologists need to weigh the risks and benefits of anticancer treatments. Patients from distant geographical areas are affected the most. Virtual management (telemedicine) has been shown to reduce health costs and improve access to care. Examples of successful technology applications include symptoms management and supervision of satellite anticancer infusion suites [7]. To our knowledge, oncologists' views about virtual management during the COVID-19 pandemic have not been described previously.

Our work demonstrates that oncologists have a high level of awareness of virtual clinics, virtual multidisciplinary teams, and virtual prescriptions (82%, 79%, and 75%, respectively). However, despite this high level of awareness, oncologists' actual involvement was significantly lower, as shown in Table 3. We presume these differences are related to the major challenges of virtual management faced by respondents, as shown in Figure 1. Challenges mentioned included a lack of physical examination (60%), patient's awareness and access (59%), a lack of physical attendance of patients (42%), IT support (37%), and the safety of virtual management (35%).

Nonetheless, we found that 46% of the surveyed oncologists responded that they "definitely" prefer to manage some cases virtually. However, only 36.0% will continue virtual management after the pandemic; we cannot explain why the proportion is low, although challenges and preferences with virtual management might be potential reasons.

No studies have previously described oncologists' views about the priority of anticancer treatments during the COVID-19 pandemic. The results of our survey demonstrated that more than 80% of oncologists preferred to continue neoadjuvant, adjuvant, and perioperative treatments. Additionally, 53% reported their preference to continue first-line palliative treatment, in contrast to 20% and 30% that preferred to interrupt second- and third-line palliative treatment, respectively. These results are consistent with Hanna et al [16], where the proposed resources utilization, allocation, and prioritization of anticancer treatments indicated a high priority for curative treatments versus a low priority for palliative treatments. Of note, this model has not been tested in clinical studies. Other studies have shown that delaying adjuvant treatments was associated with inferior survival in colon cancer [17] and breast cancer [18].

Changing the drug administration route from intravenous to oral without compromising outcomes has been reported in the literature [19]. In our study, we found that 50% and 48% of oncologists did not prefer the virtual prescription of chemotherapy and novel immunotherapy, respectively; the majority are parenteral drugs. However, 85%, 74%, and 57% of oncologists preferred the virtual prescription of hormonal therapy, bone modifying agents, and targeted therapy, respectively; the majority are oral drugs. All respondents preferred the oral route, in keeping with Hofheinz et al [19], and 53% preferred the subcutaneous route for virtual prescription. In contrast, 87% of oncologists did not prefer the intravenous route for virtual prescription.

This study has several strengths. First, we described the views of oncologists on virtual management and the priority of anticancer treatments during the COVID-19 pandemic, which has not been reported previously. Second, we used a rigorous methodology for our instrument development, validation, and administration, as no appropriate instrument previously existed. Third, we used virtual snowball sampling to identify experts in the field of oncology in the region, as there are no lists or other obvious sources for locating all practicing oncologists who are members of societies or nonmembers. It is difficult to estimate the total size of the sample.

The limitations of our study include that the number of participants in the study was relatively small, and they were mostly from Saudi Arabia. One inherent weakness of this study is its restricted participation to the Arab world, which limits the inferences that can be drawn from the data. Another important limitation is that there were differences in respondent specialties, which included medical oncology, hematology, pediatric oncology; half of the study group specialized in medical oncology. However, we tried to control for this by inviting more respondents to participate. Future research could be done with more specialties and the involvement of oncologists from different geographic regions.

Our study adds to the previous knowledge that oncologists have a high level of awareness about virtual management, although the doctors have lower actual involvement in virtual clinics, virtual multidisciplinary tumor boards, and virtual prescriptions. Our results confirm that oncologists' views on the priority of anticancer treatments are consistent with the evolving literature during the COVID-19 pandemic.

Virtual management could be implemented as an evolving method to manage a selected group of patients with cancer who live in remote locations. Mclean et al [20] showed no differences in outcomes between telehealth and usual care. This would reduce the risk associated with hospital visits and of infection transmisson. Examples of successful implementation include survivorship care, palliative care, symptoms management, and supervision of satellite anticancer infusion suites [7-9]. Similarly, this could be implemented for patients with routine follow-ups who have a low risk of relapse, as well as for patients receiving oral treatment, hormonal treatment, and bone modifying agents.

Challenges to virtual management should be addressed to improve the care of patients with cancer and to enhance oncologists' actual involvement in virtual management. This can be done by improving patients' awareness and access, improving IT support, assessing the safety of virtual management, and finding solutions to the need for physical attendance and physical examination. A number of questions remain unanswered, such as the safety of virtual management. Cancer care is complex and encompasses the need for direct doctor-patient encounters, clinical examination, medicolegal aspects, psychological support, privacy, and adequate infrastructure to support logistics. In addition, virtual management may only be feasible and applicable at some institutions. These are possible avenues for future research and will allow us to understand how these factors translate into the improvement of cancer care.

Taken together, virtual management is an evolving tool for caring for patients with cancer under certain circumstances. If it is implemented in the appropriate venues, it will improve access to care [6] and reduce the health care burden on patients with cancer. Virtual prescription of anticancer treatments during the COVID-19 pandemic has not been addressed before and is worth pursuing in further research.

Conclusions

In this regional survey, we found that oncologists have a high level of awareness about virtual management, with lower actual involvement with virtual clinics, virtual multidisciplinary tumor boards, and virtual prescriptions. Oncologists indicated that second- and third-line palliative treatments should be interrupted, while neoadjuvant, adjuvant, perioperative, and first-line palliative treatments should continue. Our results confirm that oncologists' views on the priority of anticancer treatments are consistent with the evolving literature during the COVID-19 pandemic. Challenges to virtual management should be addressed to improve the care of patients with cancer.

Acknowledgments

The authors wish to thank Roche Products, Saudi Arabia, for supporting the publication fee. We also thank all the oncologists in the MENA (Middle East and North Africa) region who participated in this survey.

Authors' Contributions

All authors contributed equally to the conception and design of the study, acquisition of data, analysis, and interpretation. All authors drafted the article and gave their final approval of the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.
[DOCX File , 16 KB - jmir_v22i6e19691_app1.docx]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
2. World Health Organization. Coronavirus disease 2019. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-06-09]
3. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3]

4. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. Lancet Oncol 2020 Mar;21(3):335-337 [FREE Full text] [doi: 10.1016/S1470-2045(20)30096-6] [Medline: 32066541]
5. Shankar A, Saini D, Roy S, Mosavi Jarrahi A, Chakraborty A, Bharti SJ, et al. Cancer Care Delivery Challenges Amidst Coronavirus Disease–19 (COVID-19) Outbreak: Specific Precautions for Cancer Patients and Cancer Care Providers to Prevent Spread. Asian Pac J Cancer Prev 2020 Mar 01;21(3):569-573. [doi: 10.31557/apjcp.2020.21.3.569]
6. Hazin R, Qaddoumi I. Teleoncology: current and future applications for improving cancer care globally. The Lancet Oncology 2010 Feb;11(2):204-210. [doi: 10.1016/s1470-2045(09)70288-8]
7. Sirintrapun SJ, Lopez AM. Telemedicine in Cancer Care. Am Soc Clin Oncol Educ Book 2018 May 23;38:540-545. [doi: 10.1200/EDBK_200141] [Medline: 30231354]
8. Al-Shamsi HO, Alhazzani W, Alhuraiji A, Coomes EA, Chemaly RF, Almuhanna M, et al. A Practical Approach to the Management of Cancer Patients During the Novel Coronavirus Disease 2019 (COVID-19) Pandemic: An International Collaborative Group. Oncologist 2020 Apr 03:2020 [FREE Full text] [doi: 10.1634/theoncologist.2020-0213] [Medline: 32243668]
9. Khairat S, Meng C, Xu Y, Edson B, Gianforcaro R. Interpreting COVID-19 and Virtual Care Trends: Cohort Study. JMIR Public Health Surveill 2020 Apr 15;6(2):e18811 [FREE Full text] [doi: 10.2196/18811] [Medline: 32252023]
10. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
11. Aday LA, Cornelius LJ. Designing and conducting health surveys: a comprehensive guide. San Francisco, CA: Jossey-Bass; 2006.
12. Argüelles Méndez L. From Lisp to FuzzyLisp. In: Studies in Fuzziness and Soft Computing. Cham: Springer; 2016:339-350.
13. Passmore C, Dobbie AE, Parchman M, Tysinger J. Guidelines for constructing a survey. Fam Med 2002 Apr;34(4):281-286. [Medline: 12017142]
14. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003 May;12(3):229-238. [doi: 10.1023/a:1023254226592] [Medline: 12769135]
15. Woodward CA. Questionnaire construction and question writing for research in medical education. Med Educ 1988 Jul;22(4):345-363. [doi: 10.1111/j.1365-2923.1988.tb00764.x] [Medline: 3173164]
16. Hanna TP, Evans GA, Booth CM. Cancer, COVID-19 and the precautionary principle: prioritizing treatment during a global pandemic. Nat Rev Clin Oncol 2020 May 2;17(5):268-270 [FREE Full text] [doi: 10.1038/s41571-020-0362-6] [Medline: 32242095]
17. Biagi JJ, Raphael MJ, Mackillop WJ, Kong W, King WD, Booth CM. Association between time to initiation of adjuvant chemotherapy and survival in colorectal cancer: a systematic review and meta-analysis. JAMA 2011 Jun 08;305(22):2335-2342. [doi: 10.1001/jama.2011.749] [Medline: 21642686]
18. Raphael MJ, Biagi JJ, Kong W, Mates M, Booth CM, Mackillop WJ. The relationship between time to initiation of adjuvant chemotherapy and survival in breast cancer: a systematic review and meta-analysis. Breast Cancer Res Treat 2016 Nov 8;160(1):17-28. [doi: 10.1007/s10549-016-3960-3] [Medline: 27632288]
19. Hofheinz R, Wenz F, Post S, Matzdorff A, Laechelt S, Hartmann JT, et al. Chemoradiotherapy with capecitabine versus fluorouracil for locally advanced rectal cancer: a randomised, multicentre, non-inferiority, phase 3 trial. Lancet Oncol 2012 Jun;13(6):579-588. [doi: 10.1016/S1470-2045(12)70116-X] [Medline: 22503032]
20. McLean S, Sheikh A, Cresswell K, Nurmatov U, Mukherjee M, Hemmi A, et al. The impact of telehealthcare on the quality and safety of care: a systematic overview. PLoS One 2013 Aug 19;8(8):e71238 [FREE Full text] [doi: 10.1371/journal.pone.0071238] [Medline: 23977001]

Abbreviations

COVID-19: coronavirus disease
IT: information technology

Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by G Capocasale, C Wu, R Ohannessian; comments to author 27.05.20; revised version received 29.05.20; accepted 04.06.20; published 24.06.20.

Please cite as:
Tashkandi E, Zeeneldin A, AlAbdulwahab A, Elemam O, Elsamany S, Jastaniah W, Abdullah S, Alfayez M, Jazieh AR, Al-Shamsi HO
Virtual Management of Patients With Cancer During the COVID-19 Pandemic: Web-Based Questionnaire Study
J Med Internet Res 2020;22(6):e19691
URL: http://www.jmir.org/2020/6/e19691/
doi:10.2196/19691
PMID:32501807

Viewpoint

Leveraging User Experience to Improve Video Consultations in a Cardiology Practice During the COVID-19 Pandemic: Initial Insights

Pieter Vandekerckhove[1], MSc; Yves Vandekerckhove[2], MD; Rene Tavernier[2], MD, PhD; Kelly De Jaegher[2], BN; Marleen de Mul[1], PhD

[1]Erasmus School of Health Policy and Management, Erasmus University, Rotterdam, Netherlands
[2]Department of Cardiology, AZ Sint-Jan Brugge, Brugge, Belgium

Corresponding Author:
Pieter Vandekerckhove, MSc
Erasmus School of Health Policy and Management
Erasmus University
Burgemeester Oudlaan 50
Rotterdam, 3000 DR
Netherlands
Phone: 31 10 408 8555
Email: vandekerckhove@eshpm.eur.nl

Abstract

During the coronavirus disease (COVID-19) pandemic, cardiologists have attempted to minimize risks to their patients by using telehealth to provide continuing care. Rapid implementation of video consultations in outpatient clinics for patients with heart disease can be challenging. We employed a design thinking tool called a customer journey to explore challenges and opportunities when using video communication software in the cardiology department of a regional hospital. Interviews were conducted with 5 patients with implanted devices, a nurse, an information technology manager and two cardiologists. Three lessons were identified based on these challenges and opportunities. Attention should be given to the ease of use of the technology, the meeting features, and the establishment of the connection between the cardiologist and the patient. Further, facilitating the role of an assistant (or virtual assistant) with the video consultation software who can manage the telehealth process may improve the success of video consultations. Employing design thinking to implement video consultations in cardiology and to further implement telehealth is crucial to build a resilient health care system that can address urgent needs beyond the COVID-19 pandemic.

(J Med Internet Res 2020;22(6):e19771) doi:10.2196/19771

KEYWORDS

telemedicine; design thinking; cardiology; patient; COVID-19; user experience

Introduction

The coronavirus disease (COVID-19) crisis has challenged health care professionals to rapidly reduce face-to-face consultations. To ensure care continuity, the use of telehealth is recommended [1]. Telehealth refers to the use of electronic services to support a broad range of remote services, such as patient care, education, and monitoring [2]. Many health systems have already invested in telehealth, and some primary care practices in the United States have appeared to adopt telehealth almost instantly [3-5].

In Belgium, cardiologists remotely triaged patients who were originally scheduled for face-to-face consultations; however, many consultations were postponed during the lockdown period. Some patients are afraid to come to the hospital due to the risk of contracting COVID-19. This increases the likelihood that patients will stay at home or postpone consultations despite deteriorating symptoms. In fact, a declining incidence of acute myocardial infarction has been witnessed in the United States during the COVID-19 pandemic [6]. However, early detection of atrial fibrillation is crucial to prevent stroke, which is a leading cause of death globally [7,8]. In addition, early detection of heart failure is necessary to prevent hospitalization and death [9].

Telehealth can help mitigate these risks by enabling continued monitoring of patients. Various telehealth solutions can be leveraged for remote cardiology monitoring [1], such as video communication software and implanted devices. The adoption of these tools is now being facilitated because financial and reimbursement restrictions are being lifted; however, further measures are needed for wider adoption of telehealth [4,10,11].

Given that rapid acceptance of telehealth during COVID-19 is critical, telehealth technologies must be easy to implement and to scale up. To achieve this, we employed design thinking [12,13] to learn from the experiences of a cardiology practice in a regional Belgian hospital where video consultations are rapidly being implemented.

Design Thinking

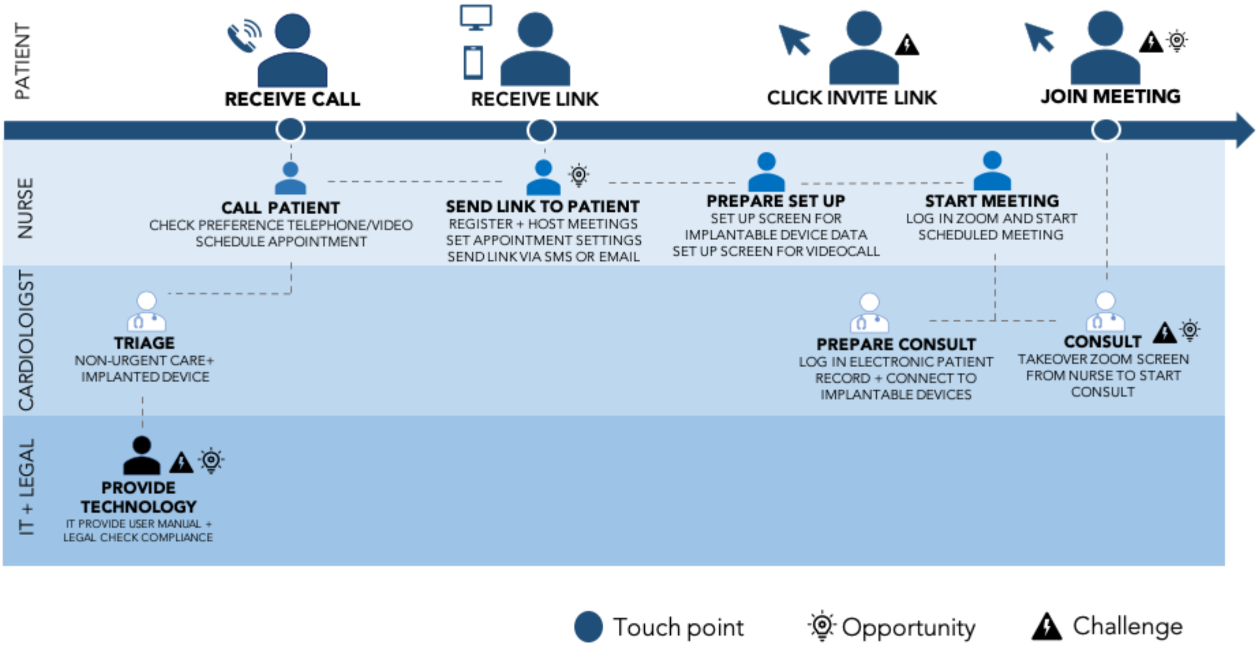
Design thinking aims to identify and solve problems in a systematic and collaborative way [12,13]. Collaborative design methods are widely used to improve electronic health (eHealth) [14], including the development of eHealth to assist heart patients with self-management [15-19]. However, design thinking research focusing on video communication software with heart patients is lacking. We used a design thinking tool called a customer journey to empathize with all stakeholders, identify the challenges facing each stakeholder, and identify opportunities to redesign the service [12]. To develop the customer journey, PV conducted telephone interviews to map the experience of the lead hospital information technology (IT) manager (who liaised with the legal team), a research nurse (KDJ), the treating cardiologist (YV), and the head of the Department of Cardiology (RT). Five patients with an implanted cardioverter-defibrillator device (ICD) were interviewed by a nurse (KDJ) shortly after the end of each video consultation about their experience. The data were analyzed in PowerPoint (Microsoft Corporation).

The treating cardiologist (YV) identified 13 patients with an ICD who were scheduled for face-to-face consultations in the outpatient clinic. The ICD patient population was prioritized because these patients could especially benefit from telehealth due to the opportunity to access their heart monitoring data remotely. These patients did not have any urgent needs, did not require a physical examination with hospital equipment, and had a telephone number on record. Of these 13 patients, 5 (38%, aged 43-64 years) were eligible for a video consultation. Of the 8 patients who were not eligible, 2 (25%) lacked a smartphone or computer, 2 (25%) were not reachable by telephone, 3 (38%) lacked the functionality for remote monitoring, and 1 (13%) had progressed to needing more urgent care.

Challenges and Opportunities for Stakeholders

The customer journey (Figure 1) shows the actions taken by stakeholders in parallel rows and the touch points (blue and white circles) where the patient is in contact with health care professionals. Three areas of challenges (triangles) and opportunities (lightbulbs) are illustrated: the lower left area is related to the provision of technology, the center area is related to inviting the patient to the consultation, and the right area relates to when the patient joins the meeting and the video consultation starts.

Figure 1. Customer journey of a video consultation for a patient with an implanted device. IT: information technology. SMS: short message service.



Video Consultation Technology: Need for Integration, Fewer Steps, and Long-Term Prospects

Available video consultation solutions were explored with the IT and legal teams based on ease of use and suitability for the patient group, short-term implementation for all stakeholders, adaptability, financial constraints, and compliance with the European Union General Data Protection Regulation (GDPR). Skype for Business (Microsoft Corporation) was accessible by all physicians, as it was installed across the hospital on desktop computers. Given its strong security, lack of additional cost, and availability in the hospital, this appeared to be a good option.

However, Skype for Business is a rigid software package that will soon be replaced by Microsoft Teams. Eventually, video communication software will be integrated into the electronic patient record software, as is already the case in larger academic hospitals in Belgium. Therefore, Skype for Business demonstrated little adaptability and no long-term prospects. Given that it would still be necessary to provide training to health care professionals and patients, other software was considered. No solution could be identified that integrated video communication software with ICD monitoring, which would suit the patients best (challenge). However, other video communication software was tested, including Zoom (Basic account, Zoom Video Communications Inc). The cardiologist was given remote training on Zoom by the researchers while they simultaneously tested its functionalities compared to Skype for Business. They considered that Zoom was easier to use, as the patient would not need to install a software package on their computer and would need to complete fewer steps to use the desktop version or smartphone app (opportunity). The IT and legal teams had heavy workloads and were only able to provide limited technical support, such as a user manual.

Receiving the Meeting Invitation: Adapting the Settings

The nurse first called each patient to make an appointment for a consultation by telephone or video. If the patient preferred a video call, they were asked if they would like to receive the call through an app on their smartphone or on a desktop computer. Video consultation appointments were made using Zoom. To reduce the workload for the cardiologist, the nurse was responsible for making appointments and was then made the host of each meeting (opportunity). Here, opportunities were identified to improve safety by changing meeting settings with a password. Other settings were changed to reduce the number of steps the patient was required to take; immediate activation of the camera was enabled when joining a meeting, bypassing the waiting room function. After these steps were completed, a link with the invitation to the video consultation was sent to the patient. Here, another opportunity was identified to circumvent email use if the patient planned to use the app by sending them the link via SMS text message. However, this required the nurse to use an anonymous mobile phone number. Patients using the desktop version of Zoom were sent an invitation link via email.

Joining Meetings: Need for a Virtual Assistant

Before the start of a video consultation, the nurse prepared two screens (opportunity). On one screen, the cardiologist could log in to the patient's device, while the other screen was prepared to show the video communication software. For all the video consultations, a backup plan was established to switch to telephone and continue the consultation on an audio-only basis. The nurse was present throughout the entire VC process to help manage problems with communication connections and record the experiences of the patient and the cardiologist.

In 2 of the 5 video consultations, the patient and the cardiologist were connected with both audio and video via Zoom. In the four cases in which a video connection was established, even without audio (replaced by audio from a telephone), the patients were positive about the experience; however, for the one case in which the video connection could not be established, the patient was disappointed. Key challenges were identified due to failures of the video or audio on either the patient's or the cardiologist's side (3 times): in one case, the cardiologist's computer ran too slowly, in another case, the patient had not signed into the Zoom meeting, and in another case, the patient could not join the Zoom meeting. In addition, the patients were not always immediately in front of their devices at the start of the meeting. The nurse had to attempt to manage these problems in the moment on both the cardiologist and patient sides. However, connections could not always be established or re-established. This reveals the need for more advanced training on the use of video communication software before adopting video consultations, for both patients and health care professionals.

Opportunities were identified to improve the audio and video connections between the patient and the cardiologist. First, the nurse could test the Zoom video and audio links with the cardiologist's and patient's devices before the start of the consultation. This would require the nurse to be the host for all scheduled video consultations before the consultations start. For example, before the first consultation started, the nurse as the host could conduct a test conversation using the cardiologist's device to determine if the audio and video were functional and if the connection with the heart device was functioning. Following this, the cardiologist could take over the account from the nurse to start the consultation. The nurse could then prepare the next scheduled patient for their video consultation by starting the next meeting as the host and checking the audio and video connections between the patient's device and the local device by holding a test conversation.

Currently, there are technical challenges in implementing this workflow, as the meeting host cannot simultaneously start and manage multiple meetings [20]. No other video communication software could be immediately identified that would overcome this obstacle. It may be possible for a chatbot or similar automated diagnostic system to help the patient navigate the steps on their own to test the audio or video connection. All these potential solutions would require additional training for the patients and health care professionals.

Lessons to Improve Video Consultations in Cardiology

We employed a design thinking tool called a customer journey, which revealed several challenges and opportunities for stakeholders in a cardiology practice when testing video consultation software. Three lessons were identified to improve the experience for stakeholders:

Ease of Using the Technology

Attempt to reduce the number of clicks or screens that must be navigated to get to a meeting and preferably avoid downloading or registration of software and activation of a microphone or video camera (these appear to be easier to manage on a smartphone). Ideally, provide the ability to access data on implanted devices and video communication software in one integrated software solution.

Meeting Features

Ensure the video consultations are secure (ie, use a password and data encryption in line with the GDPR) and facilitate the establishment of video and audio connections by automatically starting microphone and video devices for both the health care professional and the patient. Use a convenient method to send information about the meeting to the patient, such as an SMS text message or email, depending on the wishes of the patient.

Management of Video and Audio Connections

Reduce the time spent preparing and managing connectivity. One option would be for a nurse to concurrently host multiple meetings to streamline the process of switching between consultations. If this option is not available or proves to be inadequate, it will be crucial to provide additional training for patients and health care professionals. It is therefore important to select a system that will not require time-consuming training, preferably one with an automated testing system (if available).

Future Considerations

Looking toward the future, the population of heart patients in need of remote care is likely to grow given the prolongation of the COVID-19 pandemic. Using design thinking to improve telehealth for patients who are at risk of acute health problems, such as heart attack and atrial fibrillation, is therefore increasingly urgent. Remote diagnostic tools, such as remote electrocardiogram (ECG) technology, could be integrated into telehealth video communication software, as in smartphones [21]. Although there are some consumer devices on the market with up to six ECG leads [22,23], their use is still limited due to legal obstacles or financial concerns (linked to reimbursement).

Limitations

This short design thinking study was limited due to the physical constraints of COVID-19. The data collection would have benefited from more field observations. In addition, to prevent delays in the implementation process of the video consultation software, the traditional problem exploration process started immediately through testing existing software solutions. Therefore, some problems and solutions for the stakeholders may remain unexplored.

Conclusion

These initial insights highlight that even though financial regulations currently favor the use of video consultations [4,10,11], we identified many practical obstacles from a user perspective that have critical implications. Employing design thinking and involving all relevant stakeholders may help overcome these obstacles and aid further integration of telehealth and other medical device software. This further resonates with the call of the American Heart Association to conduct more human-centered research in this area [24]. In conclusion, employing design thinking to implement video consultations in cardiology and to further implement telehealth is crucial to build a resilient health care system that can address urgent needs beyond the COVID-19 pandemic.

Conflicts of Interest

None declared.

References

1. Lakkireddy DR, Chung MK, Gopinathannair R, Patton KK, Gluckman TJ, Turagam M, et al. Guidance for Cardiac Electrophysiology During the Coronavirus (COVID-19) Pandemic from the Heart Rhythm Society COVID-19 Task Force; Electrophysiology Section of the American College of Cardiology; and the Electrocardiography and Arrhythmias Committee of the Council on Clinical Cardiology, American Heart Association. Heart Rhythm 2020 Apr 01:Preprint [FREE Full text] [doi: 10.1016/j.hrthm.2020.03.028] [Medline: 32247013]
2. Schwamm LH. Telehealth: seven strategies to successfully implement disruptive technology and transform health care. Health Aff (Millwood) 2014 Feb;33(2):200-206. [doi: 10.1377/hlthaff.2013.1021] [Medline: 24493761]
3. Mehrotra A, Ray K, Brockmeyer D, Barnett M, Bender J. Rapidly Converting to "Virtual Practices?" Outpatient Care in the Era of Covid-19. NEJM Catal 2020 Apr 01:1-5. [doi: 10.1056/CAT.20.0091]
4. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care’s Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82 [FREE Full text] [doi: 10.1056/nejmp2005835]
5. Artandi M, Thomas S, Nirav S, Srinivasan M. Rapid System Transformation to More Than 75% Primary Care Video Visits within Three Weeks at Stanford: Response to Public Safety Crisis during a Pandemic. NEJM Catal 2020 Apr 21:1-11 [FREE Full text] [doi: 10.1056/CAT.20.0100]
6. Solomon MD, McNulty EJ, Rana JS, Leong TK, Lee C, Sung S, et al. The Covid-19 Pandemic and the Incidence of Acute Myocardial Infarction. N Engl J Med 2020 May 19:1-3. [doi: 10.1056/nejmc2015630]
7. Ponikowski P, Voors A, Anker S, Bueno H, Cleland J, Coats A. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016 May 20;37(27):2129-2200. [doi: 10.3410/f.718489795.793497182]
8. World Health Organization. 2016. Global Health Observatory data repository: Causes of Death Estimates for 2000-2016 URL: https://apps.who.int/gho/data/node.main.GHECOD?lang=en [accessed 2020-04-22]

9. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, ESC Scientific Document Group. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Eur Heart J 2016 Oct 07;37(38):2893-2962. [doi: 10.1093/eurheartj/ehw210] [Medline: 27567408]
10. Ohannessian R, Duong T, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
11. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
12. Doorley S, Holcomb S, Klebahn P, Segovia K, Utley J. Stanford d.school. 2018. Design Thinking Bootcamp Bootleg URL: https://s3.xopic.de/openhpi-public/courses/1NcWQVnyTA0dLYw9kHLs4e/rtfiles/35m0Q8qXYjvHO7FHuwgVgg/bootcampbootleg2010.pdf [accessed 2020-06-10]
13. Luchs M, Swan K, Griffin A. Design Thinking: New Product Development Essentials from the PDMA. Hoboken, NJ: John Wiley & Sons; Oct 07, 2015.
14. Vandekerckhove P, de Mul M, Bramer W, de Bont AA. Generative Participatory Design Methodology to Develop Electronic Health Interventions: Systematic Literature Review. J Med Internet Res 2020 Apr 27;22(4):e13780 [FREE Full text] [doi: 10.2196/13780] [Medline: 32338617]
15. Joensson K, Melholt C, Hansen J, Leth S, Spindler H, Olsen MV, et al. Listening to the patients: using participatory design in the development of a cardiac telerehabilitation web portal. Mhealth 2019;5:33 [FREE Full text] [doi: 10.21037/mhealth.2019.08.06] [Medline: 31620460]
16. Kristiansen AM, Svanholm JR, Schjødt I, Mølgaard Jensen K, Silén C, Karlgren K. Patients with heart failure as co-designers of an educational website: implications for medical education. Int J Med Educ 2017 Feb 25;8:47-58 [FREE Full text] [doi: 10.5116/ijme.5898.309e] [Medline: 28237976]
17. Woods L, Duff J, Roehrer E, Walker K, Cummings E. Design of a Consumer Mobile Health App for Heart Failure: Findings From the Nurse-Led Co-Design of Care4myHeart. JMIR Nursing 2019 Sep 23;2(1):e14633. [doi: 10.2196/14633]
18. Noergaard B, Sandvei M, Rottmann N, Johannessen H, Wiil U, Schmidt T, et al. Development of a Web-Based Health Care Intervention for Patients With Heart Disease: Lessons Learned From a Participatory Design Study. JMIR Res Protoc 2017 May 17;6(5):e75 [FREE Full text] [doi: 10.2196/resprot.7084] [Medline: 28526674]
19. Ahmed R, Toscos T, Rohani Ghahari R, Holden RJ, Martin E, Wagner S, et al. Visualization of Cardiac Implantable Electronic Device Data for Older Adults Using Participatory Design. Appl Clin Inform 2019 Sep 18;10(04):707-718. [doi: 10.1055/s-0039-1695794]
20. Zoom Help Center. Can I Host Concurrent Meetings? URL: https://support.zoom.us/hc/en-us/articles/206122046-Can-I-Host-Concurrent-Meetings- [accessed 2020-04-20]
21. Li KHC, White FA, Tipoe T, Liu T, Wong MC, Jesuthasan A, et al. The Current State of Mobile Phone Apps for Monitoring Heart Rate, Heart Rate Variability, and Atrial Fibrillation: Narrative Review. JMIR Mhealth Uhealth 2019 Feb 15;7(2):e11606 [FREE Full text] [doi: 10.2196/11606] [Medline: 30767904]
22. AliveCor. Introducing KardiaMobil 6L URL: https://www.alivecor.com/kardiamobile6l [accessed 2020-04-20]
23. Halcox JP, Wareham K, Cardew A, Gilmore M, Barry JP, Phillips C, et al. Assessment of Remote Heart Rhythm Sampling Using the AliveCor Heart Monitor to Screen for Atrial Fibrillation. Circulation 2017 Nov 07;136(19):1784-1794. [doi: 10.1161/circulationaha.117.030583]
24. Schwamm LH, Chumbler N, Brown E, Fonarow GC, Berube D, Nystrom K, et al. Recommendations for the Implementation of Telehealth in Cardiovascular and Stroke Care: A Policy Statement From the American Heart Association. Circulation 2017 Feb 14;135(7):e24-e44. [doi: 10.1161/cir.0000000000000475]

Abbreviations

COVID-19: coronavirus disease
ECG: electrocardiogram
eHealth: electronic health
GDPR: General Data Protection Regulation
ICD: implanted cardioverter-defibrillator device
IT: information technology

Edited by G Eysenbach; submitted 30.04.20; peer-reviewed by A Galpin, MA Bahrami, A Kocaballi; comments to author 23.05.20; revised version received 04.06.20; accepted 09.06.20; published 25.06.20.

Please cite as:
Vandekerckhove P, Vandekerckhove Y, Tavernier R, De Jaegher K, de Mul M
Leveraging User Experience to Improve Video Consultations in a Cardiology Practice During the COVID-19 Pandemic: Initial Insights
J Med Internet Res 2020;22(6):e19771
URL: http://www.jmir.org/2020/6/e19771/
doi:10.2196/19771
PMID:32519964

Original Paper

Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation

Viroj Tangcharoensathien[1*], PhD, MD; Neville Calleja[2*], MD, PhD; Tim Nguyen[3*], MSc; Tina Purnat[4*], MSc; Marcelo D'Agostino[5*], MSc; Sebastian Garcia-Saiso[6*], PhD; Mark Landry[7*], MSc; Arash Rashidian[8*], PhD; Clayton Hamilton[9*], MSc; Abdelhalim AbdAllah[10*], MSc; Ioana Ghiga[11*], MSc; Alexandra Hill[11*], MSc; Daniel Hougendobler[11*], MPH, JD, LLM; Judith van Andel[4*], MD; Mark Nunn[*], MSc; Ian Brooks[12*], PhD; Pier Luigi Sacco[13,14,15,16*], PhD; Manlio De Domenico[17*], PhD; Philip Mai[18*], MA, JD; Anatoliy Gruzd[18*], PhD; Alexandre Alaphilippe[19*], MSc; Sylvie Briand[11*], PhD

[1]International Health Policy Program, Ministry of Public Health, Nonthaburi, Thailand

[2]Directorate for Health Information and Research, Ministry for Health, Valetta, Malta

[3]High Impact Events Preparedness, Global Infectious Hazards Preparedness, Emergency Preparedness, Geneva, Switzerland

[4]Department of Digital Health and Innovation, Science Division, World Health Organization, Geneva, Switzerland

[5]Information Systems for Health, Evidence and Intelligence for Action in Health, Pan American Health Organization and World Health Organization Regional Office for the Americas, Washington DC, DC, United States

[6]Evidence and Intelligence for Action in Health, Pan American Health Organization and World Health Organization Regional Office for the Americas, Washington, DC, United States

[7]Department of Health Systems Development, Regional Office for South-East Asia, World Health Organization, New Delhi, India

[8]Department of Science, Information and Dissemination, Eastern Mediterranean Regional Office, World Health Organization, Cairo, Egypt

[9]Division of Health Systems and Public Health, Regional Office for Europe, World Health Organization, Copenhagen, Denmark

[10]Communications Department, Regional Office for Africa, World Health Organization, Brazzaville, Congo

[11]Global Infectious Hazards Preparedness, Emergency Preparedness, World Health Organization, Geneva, Switzerland

[12]Center for Health Informatics, School of Information Sciences, University of Illinois, Champaign, IL, United States

[13]Fondazione Bruno Kessler, Trento, Italy

[14]Venice Office, Organisation for Economic Co-operation and Development, Venice, Italy

[15]The University Institute for Modern Languages, Milan, Italy

[16]Berkman-Klein Center for Internet and Society, Harvard University, Cambridge, MA, United States

[17]Complex Multilayer Networks (CoMuNe) Research, Center for Information and Communication Technology, Bruno Kessler Foundation, Trento, Italy

[18]Ted Rogers School of Management, Ryerson University, Toronto, ON, Canada

[19]EU DisinfoLab, Brussels, Belgium

[*]all authors contributed equally

Corresponding Author:
Tina Purnat, MSc
Department of Digital Health and Innovation
Science Division
World Health Organization
20 Avenue Appia
Geneva, 1211
Switzerland
Phone: 41 22 791 2476
Email: purnatt@who.int

Related Article:

Comment in: https://www.jmir.org/2020/6/e21820/

Abstract

Background: An infodemic is an overabundance of information—some accurate and some not—that occurs during an epidemic. In a similar manner to an epidemic, it spreads between humans via digital and physical information systems. It makes it hard for people to find trustworthy sources and reliable guidance when they need it.

Objective: A World Health Organization (WHO) technical consultation on responding to the infodemic related to the coronavirus disease (COVID-19) pandemic was held, entirely online, to crowdsource suggested actions for a framework for infodemic management.

Methods: A group of policy makers, public health professionals, researchers, students, and other concerned stakeholders was joined by representatives of the media, social media platforms, various private sector organizations, and civil society to suggest and discuss actions for all parts of society, and multiple related professional and scientific disciplines, methods, and technologies. A total of 594 ideas for actions were crowdsourced online during the discussions and consolidated into suggestions for an infodemic management framework.

Results: The analysis team distilled the suggestions into a set of 50 proposed actions for a framework for managing infodemics in health emergencies. The consultation revealed six policy implications to consider. First, interventions and messages must be based on science and evidence, and must reach citizens and enable them to make informed decisions on how to protect themselves and their communities in a health emergency. Second, knowledge should be translated into actionable behavior-change messages, presented in ways that are understood by and accessible to all individuals in all parts of all societies. Third, governments should reach out to key communities to ensure their concerns and information needs are understood, tailoring advice and messages to address the audiences they represent. Fourth, to strengthen the analysis and amplification of information impact, strategic partnerships should be formed across all sectors, including but not limited to the social media and technology sectors, academia, and civil society. Fifth, health authorities should ensure that these actions are informed by reliable information that helps them understand the circulating narratives and changes in the flow of information, questions, and misinformation in communities. Sixth, following experiences to date in responding to the COVID-19 infodemic and the lessons from other disease outbreaks, infodemic management approaches should be further developed to support preparedness and response, and to inform risk mitigation, and be enhanced through data science and sociobehavioral and other research.

Conclusions: The first version of this framework proposes five action areas in which WHO Member States and actors within society can apply, according to their mandate, an infodemic management approach adapted to national contexts and practices. Responses to the COVID-19 pandemic and the related infodemic require swift, regular, systematic, and coordinated action from multiple sectors of society and government. It remains crucial that we promote trusted information and fight misinformation, thereby helping save lives.

(J Med Internet Res 2020;22(6):e19659) doi:10.2196/19659

KEYWORDS

COVID-19; infodemic; knowledge translation; message amplification; misinformation; information-seeking behavior; access to information; information literacy; communications media; internet; risk communication; evidence synthesis

Introduction

A pneumonia of unknown cause detected in Wuhan, China was first reported to the World Health Organization (WHO) Country Office in China on December 31, 2019. The disease, caused by a novel coronavirus, was subsequently named the coronavirus disease (COVID-19) and declared a Public Health Emergency of International Concern on January 30, 2020. On March 11, 2020, the WHO characterized the outbreak as a pandemic. As of June 2, 2020, more than 6.1 million cases of COVID-19 had been reported in over 200 countries and territories, resulting in more than 373,500 deaths, as reported to the WHO [1].

On February 15, 2020, the WHO Director-General Tedros Adhanom Ghebreyesus raised the concern that the epidemic was accompanied by an infodemic [2]. An infodemic is an overabundance of information—some accurate and some not—that occurs during an epidemic [3]. In a similar manner to an epidemic, it spreads between humans via digital and physical information systems. It makes it hard for people to find trustworthy sources and reliable guidance when they need it [4]. People need accurate information during epidemics to modify their behavior and protect themselves, their families, and their communities against the infection. An infodemic cannot be eliminated, but it can be managed. The management of an infodemic becomes more challenging with social media and the rapid spread of information. Similar to epidemics, the information spreads further and faster, propagated by the interconnected ways in which information is disseminated and consumed through the web, digital and social media, chat apps, TV, radio, and other communication channels.

In the context of the COVID-19 pandemic, the infodemic is exacerbated by the global scale of the emergency. Importantly, even though some misinformation may just be confusing, many false and misleading claims such as those about fake or questionable cures, or incorrect recommendations about prevention or public behavior can be harmful to life and can exacerbate the outbreak. The infodemic can be even more challenging to manage when health information messages and facts are incorporated into political narratives and online commentary that is not grounded in verified facts and evidence.

At the same time, however, with the vast amount of information related to the spread of the virus, online and offline interactions, and public opinions expressed on social media can also be valuable sources of knowledge when analyzing the dynamics of the pandemic. For instance, they can help us gauge public sentiment toward different public health measures, analyze adherence to prevention approaches, develop effective public health campaigns, track and map symptoms, predict outbreaks, and detect and combat misinformation. Big data sources and unstructured text data sets can be used alongside official and established ones for real time analytics and modeling as part of infodemic and pandemic management—although such use cases for data also raise questions related to privacy, security, and the ethics of using public and private data sets.

To counter and understand the rapidly changing landscape of the COVID-19 infodemic and develop effective strategies to mitigate its negative side effects such as the spread of misinformation, on April 7 and 8, 2020, the WHO Information Network for Epidemics (EPI-WIN) organized a 2-day global online consultation on managing the infodemic. The meeting materials, recordings of sessions, and summary illustrations of discussion are available on the WHO website [5]. This paper summarizes the proceedings and outcomes of the consultation and recommendations for further action by the WHO, its Member States, and other stakeholders.

Methods

The aim of this consultation was to crowdsource ideas to form a novel COVID-19 infodemic response framework. A 2-day online consultation with four plenary sessions and a brainstorming session was conducted entirely online using three information and communication platforms: Zoom, Slido, and Twitter. The plenary sessions used the Zoom platform for 19 talks by 21 presenters and 3 discussants. The speakers were identified based on discussion among key staff in the WHO COVID-19 infodemic management response pillar and a search of academic and grey literature for authors in key topics of interest.

Invitations to participate in the consultation were sent out through existing global and regional networks with the aim of bringing together a community of key partners already working to address the infodemic from different perspectives, including but not limited to risk communication, health information systems, research and science, policy analysis, evidence synthesis, digital health, community response, and humanitarian response. All participants were also encouraged to post their thoughts and engage with the wider audience using Twitter and the hashtag #infodemicCOVID19.

The first panel of the meeting consisted of speakers who outlined key perspectives on the infodemic from the points of view of the WHO, national health policy makers, institutes of public health, the news and media, social media platforms, the private sector, publishers of scientific journals, sociobehavioral science, and civil society. The speakers on subsequent panels presented relevant methods, tools, and evidence from past and current experiences dealing with the COVID-19 infodemic. The talks covered a wide range of topics, including risk communication and community engagement in outbreaks; fact-checking practices; identification and response strategies for misinformation; sociobehavioral science research methods; and the use of social listening, artificial intelligence (AI), and computational methods to produce insights for infodemic response decision making. The recommendations from the talks were factored into the subsequent analysis of the actions proposed by the wider group of online participants.

Only panelists and moderators were able to use audio during the Zoom webinar. In parallel, questions were submitted by participants through Slido, a question and answer as well as polling platform. Participants were able to upvote relevant questions and discuss the questions with each other by replying to the questions and comments posted on Slido. Moderators selected the three or four most upvoted questions at the end of each session to pose to panelists. Plenary sessions ran each day from 2 PM to 5 PM Central European Time (CET) to facilitate participation from a wide range of different time zones.

Slido was also used to crowdsource ideas from participants and panelists concerning the elements of a future infodemic management framework. Four breakout topics were established for which ideas could be submitted using Slido from 3 PM CET on April 7, 2020, to 12 PM CET on April 9, 2020. For the final meeting session on April 8, organizers prepared a preliminary narrative analysis of submitted ideas and organized them into subcategories of four thematic areas. For each subcategory, the most frequently named and most innovative ideas were collected for narrative categorization and summary. The selection of ideas was based on the frequency with which an idea was mentioned, the number of likes and replies it generated via Slido, and a qualitative analysis by the WHO analysis team. During the review of the submissions, similar proposals were merged into combined suggestions, and multi-thematic submissions were included in the relevant thematic areas of analysis. Two invited policy makers from the Ministry of Health of Thailand and the Ministry for Health of Malta received the preliminary summary of the narrative 90 minutes before the final session so that they could respond to it with their perspectives and comments. After the submission period for ideas closed on April 9, a final narrative analysis was conducted for all ideas submitted on the platform, including through meeting comments. That summary forms the results of this paper.

After the consultation concluded, the WHO EPI-WIN team conducted an after-action review of the meeting and disseminated lessons from the process of organizing the meeting through the WHO's internal knowledge-sharing networks. A set of follow-up actions and network-building activities was outlined so that the WHO team could engage and foster communities that contribute to the infodemic response.

Results

Meeting Participants and Discussions

A total of 1483 individuals from 111 countries and territories registered for the consultation, with 1375 and 1169 participants joining the consultation on day 1 and day 2, respectively. Registered individuals self-reported as representing the following sectors: academia or research (32%), nongovernmental organizations (21%), the private sector (17%), United Nations or intergovernmental organizations (12%), public health authorities or government (9%), health care professionals (5%), civil society (4%), and students (2%).

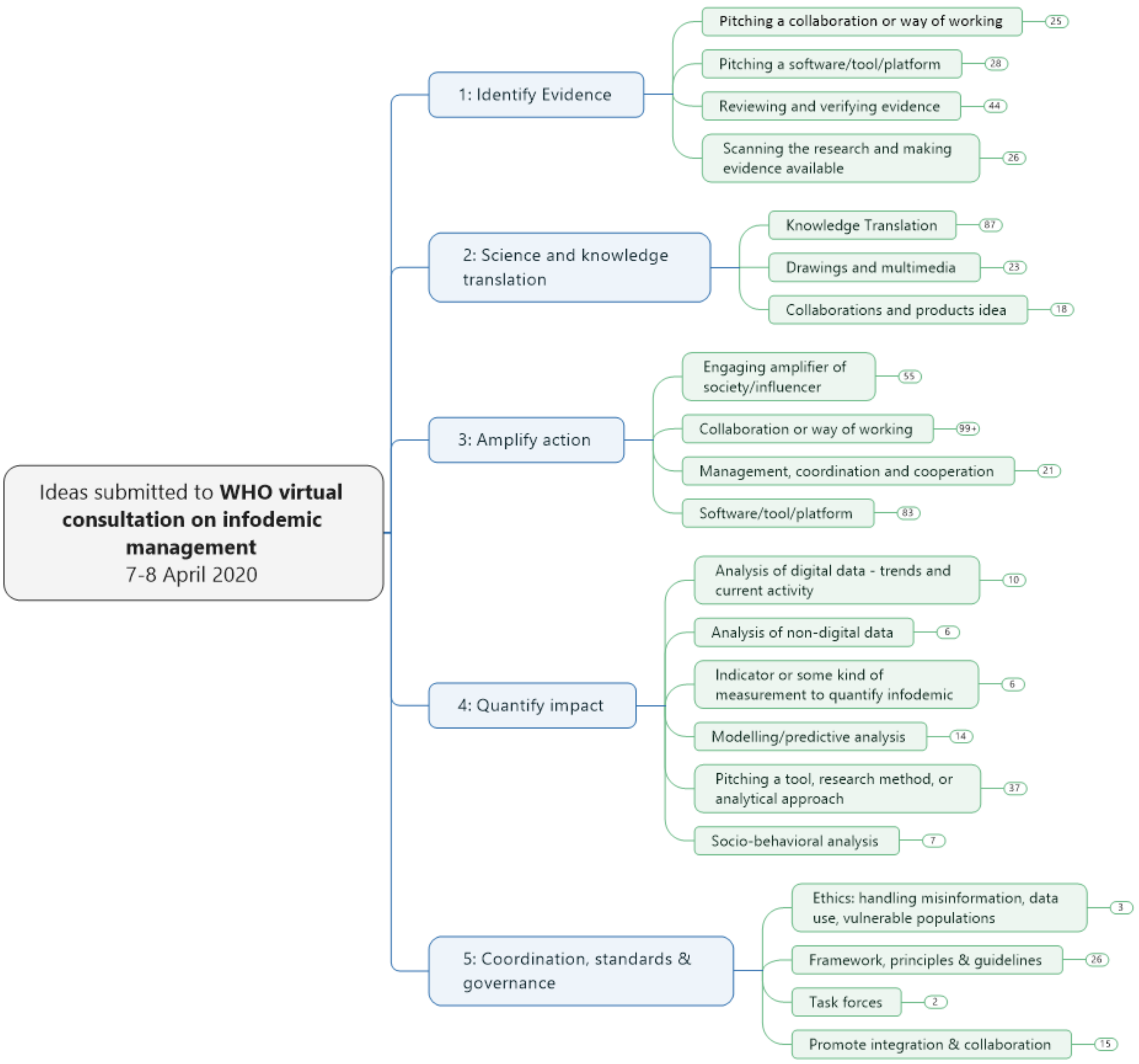
The presenters embodied a diverse, multidisciplinary array of thinking, providing the perspectives of the WHO, the media, government and policy makers, institutes of public health, scientific publishers, social media companies, academia, the private sector, civil society, and the humanitarian sector. They discussed approaches to studying the infodemic using conceptual, web-based, and sociobehavioral approaches, as well as social media analytics, and they demonstrated various tools and techniques designed to check facts and measure and respond to the infodemic. The discussions were grounded in lessons from the speakers' past experiences and research.

Both the presentations and the follow-up emphasized that we are living in an interconnected world, in which the infodemic recognizes no boundaries. Information is transmitted and shared worldwide, traveling fast through social media apps and platforms, online forums, news sites, television, radio, and many other channels. Citizens exchange information more quickly than ever before, collectively experiencing in their own everyday lives the effects and changes brought about by the pandemic and the actions undertaken to respond to it. At the same time, those supporting the response are also experiencing the pandemic and infodemic through their families and communities, as well as in their professional work and networks.

Analysis of Suggested Actions

Using a narrative analysis, we further grouped the submitted ideas into a set of five COVID-19 infodemic management areas: (1) scanning and verifying evidence (18%); (2) explaining the science (20%); (3) amplifying the reach of messages (44%); (4) measuring the infodemic and assessing trends and impacts (12%); and (5) coordination and governance (6%). The 594 collected suggestions were further summarized by theme (Figure 1).

Figure 1. Summary categorization and count of ideas submitted by the end of the consultation period. WHO: World Health Organization.



Almost half of the proposed actions dealt with how to amplify the reach of credible messages. They suggested specific actions that the WHO or Member States could undertake, such as providing coordination and development of guidelines and frameworks; building coalitions with UN agencies, fact-checking organizations, data scientists, the AI community, social media companies, and journalists; and generating specific communication strategies and tools to reach all communities and vulnerable groups. In addition, many called for specific practical actions such as tailoring messages to specific audiences, mindful of context and literacy; developing dashboards to monitor the infodemic; developing and applying research methods to understand the infodemic at the level of information flows, populations, individuals, and communities; and analyzing the adherence to, and impact of, public health measures.

The recommendations and ideas submitted indicate that we lean on existing infodemic experience, knowledge, and tools from a range of sources, including lessons from previous outbreaks such as Zika, severe acute respiratory syndrome, and Ebola; studies on misinformation around vaccine safety and methods for addressing vaccine communication, trust, and misinformation; existing tools for reaching communities with lower internet use or literacy; and experience in community outreach and humanitarian action. At the same time, new tools were suggested to assist the process of reviewing information and taking action, and to simplify the generation and dissemination of information, messages, and materials. It was suggested that we use multidisciplinary approaches in research and analysis, such as AI for natural language processing or treatment of unstructured and text data, as well as sociobehavioral, ethnographic, and medical anthropology research methods. The submissions also called for cooperation and coordination between governments, public health authorities, journalists, fact checkers, communities, social media, the private sector, academia, and others. A detailed mind map of suggestions for action is available in Multimedia Appendix 1.

Policy Implications

The analysis team further distilled the results of the consultation into a set of 50 proposed actions for a framework for managing infodemics in health emergencies (see Multimedia Appendix 2). The consultation revealed six implications for governments and policy makers to consider.

First, interventions and messages must be based on science and evidence, and must reach citizens and enable them to make informed decisions on how to protect themselves and their communities in a health emergency [6,7]. These basic foundations of emergency preparedness and response have been challenged and adapted in the current pandemic context by the rapid exchange of information and consequent shortened time frames for decision making.

To be able to provide policy solutions to the challenges caused by the infodemic, we must acknowledge the challenges that exist in managing the creation and dissemination of trusted information so that it is not excessive, overwhelming, or confusing and deciding when and how to counter misinformation.

- Work is required to slow down and streamline the flow of information of all kinds. This should be guided by a unified strategy for producing and disseminating trusted information, and a constantly updated set of guidelines to counter and correct misinformation using a measured approach firmly grounded in state-of-the-art sociobehavioral research. COVID-19 has resulted in an explosion of evidence generation and synthesis activities—“not only an ‘infodemic on top of the pandemic,’ but also an ‘infodemic on top of the infodemic’” [8]. These activities should be internationally coordinated to avoid duplication while guaranteeing epistemic diversity.
- As the body of evidence grows during an emergency, guidelines and messages may change or be amended. Care needs to be taken to ensure propagation of up-to-date information to communities and individuals through all appropriate channels, including to those communities for which web-based sources are not the primary source of information and those for which information needs to be adapted to local languages, literacy levels, and contexts.
- In parallel, scientific findings must be collated, reviewed, appraised, and assessed for relevance to help form recommendations and policies that have the greatest possible positive impact on the health of individuals and populations [9]. Scientific and public health institutions have central roles to play in this process. Editors of medical journals could also help manage the infodemic problem by, for example, producing plain language summaries for journalists and the public to accompany each article related to COVID-19.
- The capacity must be in place to promote trusted content; check facts; and monitor, verify, report, and respond to misinformation. Work is required to verify and counter the spread of misinformation and introduce effective regulatory approaches to limit its impact. This could be strengthened through partnerships between public health authorities, communities, journalists, fact-checking initiatives, social media platforms, and other communication channels.

Second, knowledge should be translated into actionable behavior-change messages, presented in ways that are understood by and accessible to all individuals in all parts of all societies. Cultural and contextual sensitivity in the platforms and messages used, and translation into local languages are essential. Coordinated work and partnering with a variety of stakeholders, including civil society, is required to ensure the availability of information targeted at vulnerable or hard-to-reach communities via nondigital routes. An on-the-ground network of global field workers could help reach out to highly vulnerable people to ensure they can access reliable information, as many citizens around the world still do not have access to pandemic information on the internet.

Third, governments should reach out to key communities to ensure their concerns and information needs are understood, tailoring advice and messages to address the audiences they represent. Through this process, communities of all kinds, whether neighborhood, religious, professional, or otherwise, should amplify the right public health messages in ways that are user-friendly and can lead to the right changes in behavior. For example, active engagement calls and dialogue could be established between public health authorities and private sector employers, telecom companies, the food and agriculture sector, faith-based organizations, health care and medical professional associations, and the media.

Fourth, to strengthen the analysis and amplification of information impact, strategic partnerships should be formed across all sectors, including but not limited to the social media and technology sectors, academia, and civil society. Through strategic partnerships with health authorities, these platforms can place and prioritize relevant information and advice, ensuring it is seen by citizens, helping to fact-check, measuring and describing the infodemic, tracking trends, and observing and analyzing the impact of messages and interventions on population behavior. There is a wealth of information on these platforms that can help us to understand the sentiments of different populations and guide effective public health measures. Rather than a complement to public health, the infodemic dimension should be considered a pillar of an integrative approach to public health in complex knowledge societies.

Fifth, health authorities should ensure that these actions are informed by reliable information that helps them understand the circulating narratives and changes in the flow of information, questions, and misinformation in communities. Analysis of online conversations; narratives; and TV, radio, and news media could be systematically applied and paired with fact-checking resources. Analysis of circulating narratives, knowledge, risk perceptions, behaviors, and trust at population and community levels could provide rapid snapshots to inform appropriate policy interventions. Mixed-methods research can monitor trends in acceptance of public health measures. Examples of such research methods include sociobehavioral research and sentiment and media analysis of digital information from online conversations; TV, radio, and news media; and community dipstick surveys.

Sixth, following experiences to date in responding to the COVID-19 infodemic and the lessons from other disease outbreaks, infodemic management approaches should be further developed to support preparedness and response, and to inform risk mitigation, and enhanced through data science and sociobehavioral and other research.

Discussion

This WHO technical consultation was entirely held online, with no panelists or participants travelling to the meeting. In terms of the number of participants, it was one of the biggest meetings ever convened by the WHO—comparable to the annual World Health Assembly. It was also the first consultation to address the phenomenon of infodemic management in health emergencies. This online consultation was an effective and cost-efficient way of reaching and interacting with a large and diverse community and producing solid outcomes.

The infodemic impacts citizens in every country, and addressing it is a new and centrally important challenge in responding to the COVID-19 pandemic—and will be so for future pandemics. The consultation discussions and online brainstorming generated clear themes that can inform discussions and actions for effective management of infodemics in all countries. The participants, presenters, and experts collectively agreed that, today more than ever, it is crucially important for authorities, other stakeholders, and the public to have access to the right data and information, at the right time, and in the right formats. The better the data and information available at all levels, the smarter and more effective the response to the pandemic will be [10].

The extraordinary interdependence of the sociobehavioral dimension of the pandemic with individual and public health makes the infodemic a serious threat but is also an opportunity to shape views and behaviors in our societies. The spread of misinformation [11,12] can trigger behavioral responses that, in turn, can further expose individuals and communities to health risks. Although the impact of misinformation on society is still under scrutiny, there is increasingly convincing evidence that deliberate misinformation operations and social manipulation [13] have taken place during major events in the past—including in several high profile political instances [14]—and that they might play an important role during other high profile, strategically important or exceptional events, including the COVID-19 pandemic.

In the study of epidemics, epidemiology experienced a turning point during the 20th century through the application of mathematical and statistical methods to this scientific field. In the study of infodemics, the 21st century will see the development of infodemiology, a novel scientific discipline required to unravel the complex propagation patterns of the infodemic. In practice, infodemiology requires a transdisciplinary approach integrating applied mathematics, social, and behavioral sciences; communication and information science; digital health research; data science; and complexity science. This scientific discipline, as it addresses research priorities driven by health policy-making needs, can generate evidence to inform the development of tools, methods, and infodemic management interventions, and contribute to the monitoring of public health interventions during outbreaks, thus strengthening outbreak preparedness and response in health emergencies.

Responses to the COVID-19 pandemic and the related infodemic require swift, regular, systematic, and coordinated action from multiple sectors of society and government. It remains crucial that we promote trusted information and fight misinformation, thereby contributing to saving lives as the pandemic continues to unfold and run its course. This requires timely translation of evidence into knowledge that people can use, adapted to their local cultures, languages, and contexts. This needs to be supported by facts and analytics, backed up with constant monitoring of the impact of trusted information, and work to counteract misinformation.

We call on citizens from all parts of society to demand reliable, evidence-based information, and to take actions that empower their communities to use trusted information to protect the most vulnerable and themselves.

Acknowledgments

The authors would like to thank all participants for contributing their ideas and opinions during the event, through Twitter, online polls, and chats, and the EPI-WIN team and Olga Fradkina, for contributing to the smooth organization of the event and its accompanying activities. The authors are grateful for the help provided by the reviewers Myrna Marti and Felipe Mejia Medina.

Tim Nguyen, Tina Purnat, Marcelo d'Agostino, Sebastian Garcia-Saiso, Mark Landry, Arash Rashidian, Clayton Hamilton, Abdelhalim AbdAllah, Ioana Ghiga, Alexandra Hill, Daniel Hougendobler, Judith Van Andel, and Sylvie Briand are staff members of the World Health Organization. The authors alone are responsible for the views expressed in this paper and they do not present the views of the organization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
An interactive mind map of suggested actions submitted during the consultation.
[ZIP File (Zip Archive), 1401 KB - jmir_v22i6e19659_app1.zip]

Multimedia Appendix 2
A set of 50 actions for an infodemic management framework that have been consolidated from the suggestions received during consultation.
[PDF File (Adobe PDF File), 186 KB - jmir_v22i6e19659_app2.pdf]

References

1. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-06-02]
2. Munich Security Conference. World Health Organization. 2020 Feb 15. URL: https://www.who.int/dg/speeches/detail/munich-security-conference [accessed 2020-04-26]
3. Coronavirus disease 2019 (COVID-19) situation report – 86. World Health Organization. 2020 Apr 15. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200415-sitrep-86-covid-19.pdf?sfvrsn=c615ea20_6 [accessed 2020-04-26]
4. World Health Organization. Infodemic management: a key component of the COVID-19 global response. Weekly Epidemiological Record 2020;95(16):145-160 [FREE Full text]
5. WHO ad-hoc online consultation on managing the COVID-19 infodemic. World Health Organization. URL: https://www.who.int/teams/risk-communication/who-ad-hoc-online-consultation-on-managing-the-covid-19-infodemic [accessed 2020-06-02]
6. Risk communication and community engagement readiness and response to coronavirus disease (COVID-19): interim guidance, 19 March 2020. World Health Organization. URL: https://www.who.int/publications-detail/risk-communication-and-community-engagement-readiness-and-initial-response-for-novel-coronaviruses-(-ncov)
7. WHO outbreak communication guidelines. World Health Organization. 2005. URL: https://www.who.int/csr/resources/publications/WHO_CDS_2005_28/en/
8. Petras R. Supporting the fight against COVID-19 infodemic in SE Asia. WHO Ad Hoc Consultation on the COVID-19 Infodemic Management Framework.: World Health Organization URL: https://www.who.int/teams/risk-communication/who-ad-hoc-online-consultation-on-managing-the-covid-19-infodemic
9. WHO handbook for guideline development, 2nd edition. World Health Organization. 2014. URL: https://apps.who.int/iris/handle/10665/145714
10. COVID-19 factsheets: why data disaggregation is key during a pandemic. Pan American Health Organization. 2020 Apr 13. URL: https://iris.paho.org/handle/10665.2/52002
11. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]
12. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
13. Shao C, Ciampaglia GL, Varol O, Yang K, Flammini A, Menczer F. The spread of low-credibility content by social bots. Nat Commun 2018 Nov 20;9(1):4787 [FREE Full text] [doi: 10.1038/s41467-018-06930-7] [Medline: 30459415]
14. Stella M, Ferrara E, De Domenico M. Bots increase exposure to negative and inflammatory content in online social systems. Proc Natl Acad Sci U S A 2018 Dec 04;115(49):12435-12440 [FREE Full text] [doi: 10.1073/pnas.1803470115] [Medline: 30459270]

Abbreviations

AI: artificial intelligence
CET: Central European Time
COVID-19: coronavirus disease
EPI-WIN: WHO Information Network for Epidemics
WHO: World Health Organization

Edited by G Eysenbach; submitted 08.06.20; peer-reviewed by K Reuter; accepted 18.06.20; published 26.06.20.

Please cite as:
Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D’Agostino M, Garcia-Saiso S, Landry M, Rashidian A, Hamilton C, AbdAllah A, Ghiga I, Hill A, Hougendobler D, van Andel J, Nunn M, Brooks I, Sacco PL, De Domenico M, Mai P, Gruzd A, Alaphilippe A, Briand S
Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation
J Med Internet Res 2020;22(6):e19659
URL: http://www.jmir.org/2020/6/e19659/
doi:10.2196/19659
PMID:32558655

Original Paper

COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep Learning Framework With a Single Chest CT Image: Model Development and Validation

Hoon Ko[1*], MSc; Heewon Chung[1*], MSc; Wu Seong Kang[2*], MD, PhD; Kyung Won Kim[3], MD; Youngbin Shin[3], MSc; Seung Ji Kang[4], MD; Jae Hoon Lee[5], MD; Young Jun Kim[5], MD; Nan Yeol Kim[2], MD, PhD; Hyunseok Jung[6], MD; Jinseok Lee[1], PhD

[1]Department of Biomedical Engineering, Wonkwang University College of Medicine, Iksan-si, Republic of Korea

[2]Department of Trauma Surgery, Wonkwang University Hospital, Iksan-si, Republic of Korea

[3]Department of Radiology and Research Institute of Radiology, Asan Image Metrics, Clinical Trial Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[4]Department of Internal Medicine, Chonnam National University Medical School, Gwangju-si, Republic of Korea

[5]Department of Internal Medicine, Wonkwang University Hospital, Iksan-si, Republic of Korea

[6]Department of Radiology, Wonkwang University Hospital, Iksan-si, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jinseok Lee, PhD
Department of Biomedical Engineering
Wonkwang University College of Medicine
460 Iksan-daero
Iksan-si, 54538
Republic of Korea
Phone: 82 63 850 5114
Email: gonasago@wku.ac.kr

Abstract

Background: Coronavirus disease (COVID-19) has spread explosively worldwide since the beginning of 2020. According to a multinational consensus statement from the Fleischner Society, computed tomography (CT) is a relevant screening tool due to its higher sensitivity for detecting early pneumonic changes. However, physicians are extremely occupied fighting COVID-19 in this era of worldwide crisis. Thus, it is crucial to accelerate the development of an artificial intelligence (AI) diagnostic tool to support physicians.

Objective: We aimed to rapidly develop an AI technique to diagnose COVID-19 pneumonia in CT images and differentiate it from non–COVID-19 pneumonia and nonpneumonia diseases.

Methods: A simple 2D deep learning framework, named the fast-track COVID-19 classification network (FCONet), was developed to diagnose COVID-19 pneumonia based on a single chest CT image. FCONet was developed by transfer learning using one of four state-of-the-art pretrained deep learning models (VGG16, ResNet-50, Inception-v3, or Xception) as a backbone. For training and testing of FCONet, we collected 3993 chest CT images of patients with COVID-19 pneumonia, other pneumonia, and nonpneumonia diseases from Wonkwang University Hospital, Chonnam National University Hospital, and the Italian Society of Medical and Interventional Radiology public database. These CT images were split into a training set and a testing set at a ratio of 8:2. For the testing data set, the diagnostic performance of the four pretrained FCONet models to diagnose COVID-19 pneumonia was compared. In addition, we tested the FCONet models on an external testing data set extracted from embedded low-quality chest CT images of COVID-19 pneumonia in recently published papers.

Results: Among the four pretrained models of FCONet, ResNet-50 showed excellent diagnostic performance (sensitivity 99.58%, specificity 100.00%, and accuracy 99.87%) and outperformed the other three pretrained models in the testing data set. In the additional external testing data set using low-quality CT images, the detection accuracy of the ResNet-50 model was the highest (96.97%), followed by Xception, Inception-v3, and VGG16 (90.71%, 89.38%, and 87.12%, respectively).

Conclusions: FCONet, a simple 2D deep learning framework based on a single chest CT image, provides excellent diagnostic performance in detecting COVID-19 pneumonia. Based on our testing data set, the FCONet model based on ResNet-50 appears to be the best model, as it outperformed other FCONet models based on VGG16, Xception, and Inception-v3.

(J Med Internet Res 2020;22(6):e19569) doi:10.2196/19569

KEYWORDS

COVID-19; deep learning; convolutional neural networks, transfer learning; chest CT; CT; neural network; pneumonia; artificial intelligence; diagnosis; pneumonia; scan

Introduction

The coronavirus disease (COVID-19) pandemic is currently a global health crisis; more than 1,700,000 cases had been confirmed worldwide and more than 100,000 deaths had occurred at the time of writing this paper [1]. COVID-19, an infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is highly contagious and has spread rapidly worldwide. In severe cases, COVID-19 can lead to acute respiratory distress, multiple organ failure, and eventually death. Countries are racing to slow the spread of the virus by testing and treating patients in the early stage as well as quarantining people who are at high risk of exposure due to close contact with patients with confirmed infection. In addition, early diagnosis and aggressive treatment are crucial to saving the lives of patients with confirmed infection [2].

COVID-19 is typically confirmed by viral nucleic acid detection using reverse transcription–polymerase chain reaction (RT-PCR) [3]. However, the sensitivity of RT-PCR may not be sufficiently high; it ranges from 37% to 71% according to early reports [4-6]. Thus, RT-PCR can afford a substantial number of false negative results due to inadequate specimen collection, improper extraction of nucleic acid from the specimen, or collection at a too-early stage of infection. A chest computed tomography (CT) scan can be used as an important tool to diagnose COVID-19 in cases with false negative results by RT-PCR [6-9].

Recently, a multinational consensus statement from the Fleischner Society was issued to guide chest imaging during the COVID-19 pandemic in different clinical settings [6]. According to this consensus statement, in a setting such as South Korea, where detecting patients at an early stage and isolating all patients and people with high risk of exposure is essential, CT is a relevant screening tool due to its greater sensitivity for detecting early pneumonic changes. CT can also contribute to the management and triage of the disease by detecting severe cases. In addition, chest CT is noninvasive and is easy to perform in an equipped facility.

However, radiologic diagnostic support is not maintained 24 hours per day in many institutions [10]. In addition, CT may show similar imaging features between COVID-19 and other types of pneumonia, thus hampering correct diagnosis by radiologists. Indeed, in a study that evaluated radiologists' performance in differentiating COVID-19 from other viral pneumonia, the median values and ranges of sensitivity and specificity were 83% (67%-97%) and 96.5% (7%-100%), respectively [11].

The use of artificial intelligence (AI) may help overcome these issues, as AI can help maintain diagnostic radiology support in real time and with increased sensitivity [8,12]. In this era of worldwide crisis, it is crucial to accelerate the development of AI techniques to detect COVID-19 and to differentiate it from non–COVID-19 pneumonia and nonpneumonia diseases in CT images. Therefore, we aimed to rapidly develop an AI technique using all available CT images from our institution as well as publicly available data.

Methods

Data Sets and Imaging Protocol

This study was approved by the institutional review boards of Wonkwang University Hospital (WKUH) and Chonnam National University Hospital (CNUH). Informed consent was waived. Table 1 summarizes the training, testing, and additional validation data sets. In this study, we initially collected data from 3993 chest CT images, which were categorized into COVID-19, other pneumonia, and nonpneumonia disease groups.

For the COVID-19 data group, we used a total of 1194 chest CT images: 673 chest CT images (56.3%, from 13 patients) from CNUH, 421 images (35.3%, from 7 patients) from WKUH, and 100 images (83.8%, 60 patients) from the Italian Society of Medical and Interventional Radiology (SIRM) public database [13]. The 20 patients from CNUH and WKUH included 9 male patients and 11 female patients, with an average age of 59.6 years (SD 17.2). Regarding the COVID-19 data from WKUH and CNUH, all the patients with COVID-19 tested positive for the virus by RT-PCR viral detection, and the CT images were acquired between December 31, 2019 and March 25, 2020. The median period from symptom onset to the first chest CT examination was 8 days (range 2-20 days). The most common symptoms were fever (75%) and myalgia (30%). In addition, according to previous studies related to COVID-19 by Zhao's group [14] from January 19 and March 25, 2020, 264 low-quality chest CT images were used as additional testing data. In summary, 1194 COVID-19 images (80 patients) from WKUH, CNUH, and SRIM were split into the training data set (955 images, 80.0%) and testing data set (239 images, 20.0%). For the additional testing, 264 COVID-19 images (264 patients) from the low-quality image data set were used.

Table 1. Summary of the training, testing, and additional testing data sets (N=4257).

Data type, data source, and data group	Training images, n (%)	Testing images, n (%)
Training and testing data		
WKUH[a]		
COVID-19 pneumonia (n=421)	337 (80.0)	84 (20.0)
Other pneumonia (n=1357)	1086 (80.0)	271 (20.0)
Nonpneumonia and normal lung (n=998)	798 (80.0)	200 (20.0)
Lung cancer (n=444)	355 (80.0)	89 (20.0)
CNUH[b]		
COVID-19 pneumonia (n=673)	538 (80.0)	135 (20.0)
SIRM[c]		
COVID-19 pneumonia (n=100)	80 (80.0)	20 (20.0)
Additional external testing data		
Low-quality CT images from papers		
COVID-19[d] pneumonia (n=264)	0 (0.0)	264 (100.0)

[a] WKUH: Wonkwang University Hospital.

[b] CNUH: Chonnam National University Hospital.

[C] SIRM: Italian Society of Medical and Interventional Radiology.

[d] COVID-19: coronavirus disease.

For the other pneumonia data group, we selected 1357 chest CT images from 100 patients diagnosed with non–COVID-19 pneumonia at WKUH between September 1, 2019, and March 30, 2020. The average age of this group was 62.5 years (SD 17.2), with 68 male and 32 female patients. For the nonpneumonia data group, we also selected 1442 chest CT images from 126 patients who had no lung parenchymal disease or lung cancers at WKUH between January 2009 and December 2014. The average age of these patients was 47 years (SD 17), with 63 male patients (721/1442 images, 50.0%) and 63 female patients (721/1442 images, 50.0%). The patient demographic statistics of the COVID-19 and other pneumonia groups are summarized in Table 2. In this table, other pneumonia (not COVID-19) was categorized into two different types based on clinical characteristics: 68 cases of community-acquired pneumonia (onset 48 hours before hospital admission) and 32 cases of hospital-acquired pneumonia (onset 48-72 hours after hospital admission). Of these other pneumonia patients, 24/100 (24.0%) received laboratory confirmation of the etiology, 21 (21.0%) were confirmed to be bacterial culture positive, 3 (3.0%) were viral influenza positive by RT-PCR, and 76 (76.0%) were negative. Regarding the imaging protocols, each volumetric examination contained approximately 51 to 1094 CT images, with varying slice thicknesses from 0.5 millimeters to 3 mm. The reconstruction matrix was 512×512 pixels, with in-plane pixel spatial resolution from 0.29×0.29 to 0.98×0.98 square millimeters.

The data from WKUH, CNUH, and SIRM were randomly split with a ratio of 8:2 into a training set and a testing set, respectively, in a stratified fashion. In addition, the data for each group (WKUH, CNUH, and SIRM) were spread over different splits with a ratio of 8:2. The training data set was then further separated into sets used for training the model (80% of the training set) and for internal validation (20% of the training set). The testing set was used only for independent testing of the developed models and was never used for training the model or for internal validation. Furthermore, we tested the trained model with the additional external validation data set of low-quality images to evaluate the external generalizability of the model.

Table 2. Demographic data of patients with COVID-19 and other pneumonia.

Characteristic	COVID-19[a] pneumonia (n=20)	Other pneumonia (n=100)	*P* value
Age (years), mean (SD)	59.6 (17.2)	60.1 (17.1)	.91
Male sex, n (%)	9 (45.0)	68 (68.0)	.12
Community-acquired pneumonia, n (%)	20 (100.0)	68 (68.0)	.007
Hospital-acquired pneumonia, n (%)	0 (0.0)	32 (32.0)	
Microbiological study, n (%)			
COVID-19 positive (RT-PCR[b])	20 (100.0)	0 (0.0)	<.001
Other virus positive (influenza)	0 (0.0)	3 (3.0)	
Bacterial culture positive	0 (0.0)	21 (21.0)	
Unknown	0 (0.0)	76 (76.0)	
Human radiologist's diagnosis, n (%)			
Atypical pneumonia or viral pneumonia	20 (100.0)	15 (15.0)	N/A[c]
Pneumonia	0 (0.0)	77 (77.0)	
Aspiration pneumonia	0 (0.0)	26 (26.0)	
Necrotizing pneumonia	0 (0.0)	5 (5.0)	
Tuberculosis	0 (0.0)	5 (5.0)	
Empyema	0 (0.0)	3 (3.0)	
Emphysema	0 (0.0)	9 (9.0)	
Bronchiectasis	0 (0.0)	4 (4.0)	
Interstitial lung disease	0 (0.0)	1 (1.0)	

[a] COVID-19: coronavirus disease.

[b] RT-PCR: reverse transcription–polymerase chain reaction.

[c] N/A: not applicable.

Preprocessing

For the data acquired from WKUH and CNUH, we converted Digital Imaging and Communications in Medicine (DICOM) images to one-channel grayscale PNG images to standardize the image file format, as the images in the low-quality image data set were in PNG format. To minimize the information loss, we first displayed the DICOM images using a lung window with a 1500 Hounsfield unit window width and a –600 HU window level [15,16] and converted the images to PNG format. Subsequently, we rescaled the images to a size of 256×256 pixels and normalized the pixel values to a range between 0 and 1. All of the converted PNG format images were confirmed by three radiologists to determine any loss of image information related to pulmonary diseases. For the data from SIRM, the original JPEG format was also reformatted to the PNG format, and the images were rescaled and normalized in the same manner. For the low-quality image data set, we also rescaled and normalized the images. In this study, no further preprocessing such as lung segmentation was performed.

Image Augmentation

To reduce overfitting of the training image data, we employed two distinct forms of data augmentation: image rotation and zoom. In the data augmentation method for the rotation, angles of rotation between –10° and 10° were randomly selected. Regarding the zoom, the range was randomly selected between 90% (zoom-in) and 110% (zoom-out). Either rotation or zoom was randomly selected 10 times for each training image. By applying data augmentation, we increased the number of images in the training data set to 31,940. Table 3 shows the number of augmented images for training in each group.

Table 3. Augmented images for training in each group (N=31,940).

Data source and group	Augmented images for training, n (%)
WKUHx[a]	
COVID-19[b] pneumonia	3370 (10.6)
Other pneumonia	10,860 (34.0)
Nonpneumonia and normal lung	7890 (24.7)
Lung cancer	3550 (11.1)
CNUH[c]	
COVID-19 pneumonia	5380 (16.8)
SIRM[d]	
COVID-19 pneumonia	800 (2.5)

[a] WKUH: Wonkwang University Hospital.

[b] COVID-19: coronavirus disease

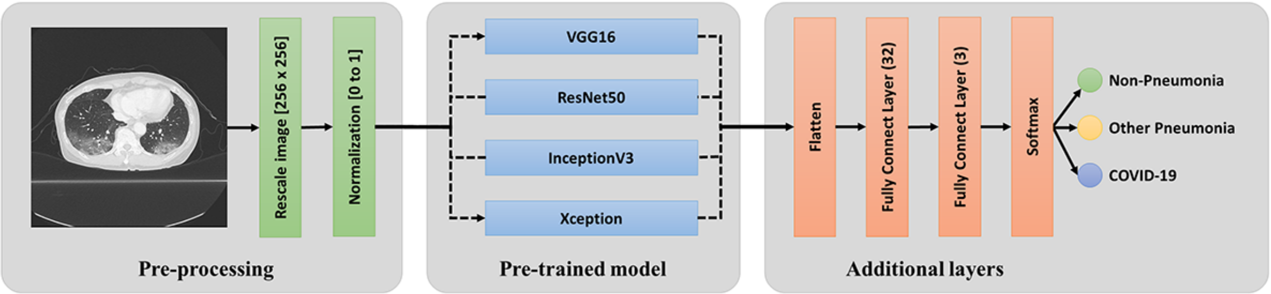
[c] CNUH: Chonnam National University Hospital.

[d] SIRM: Italian Society of Medical and Interventional Radiology.

The Fast-Track COVID-19 Classification Network for COVID-19 Classification

We developed a simple 2D deep learning framework based on a single chest CT image for the classification of COVID-19 pneumonia, other pneumonia, and nonpneumonia, named the fast-track COVID-19 classification network (FCONet; Figure 1). FCONet was developed by transfer learning based on one of the following four pretrained convolutional neural network (CNN) models as a backbone: VGG16 [17], ResNet-50 [18], Inception-v3 [19], and Xception [20]. Transfer learning is a popular method in computer vision because it enables an accurate model to be built in a short time [21]. With transfer learning, instead of starting the learning process from an optimal model search, one can start it from patterns that were learned when solving a different problem. To minimize the training time, we initially used the predefined weights for each CNN architecture, which were further updated through the learning process of classification of COVID-19 pneumonia, other pneumonia, and nonpneumonia.

Figure 1. Scheme of FCONet, a 2D deep learning framework based on a single chest CT image for the classification of COVID-19 pneumonia, other pneumonia, and non-pneumonia. COVID-19: coronavirus disease.



Input Layer

After the simple preprocessing stage, in the *input layer*, we arranged three channels (256×256×3 pixels) by copying the one-channel normalized image. The three-channel images were fed into *the pretrained model layers*.

Pretrained Model Layers

A pretrained model is a model that was trained on a large benchmark data set to solve a similar problem to the one we want to solve. In the *pretrained model layers*, we included one of the four pretrained models (VGG16, ResNet-50, Inception-v3, and Xception). Each model comprises two parts: a convolutional base and a classifier. The convolutional base is composed of a stack of convolutional and pooling layers to generate features from the images. The role of the classifier is to categorize the image based on the extracted features. In our pretrained model layers, we retained the convolutional base and removed the classifier, which was replaced by another classifier for COVID-19, other pneumonia, or nonpneumonia.

Additional Layers

The activations from the *pretrained model layers* were fed into the *additional layers*. The layers acted as classifiers for COVID-19 pneumonia, other pneumonia, and nonpneumonia. In the additional layers, we first flattened the activations and connected two fully connected layers; one of the layers consisted of 32 nodes, and the other consisted of three nodes. Subsequently, the three activations from the second fully connected layer were fed into a SoftMax layer, which provided the probability for each of COVID-19, other pneumonia, and nonpneumonia.

Implementation

We implemented FCONet using the TensorFlow package, which provides a Python application programming interface (API) for tensor manipulation. We also used Keras as the official front end of TensorFlow. We trained the models with the Adam optimizer [22] and the categorical cross-entropy cost function with a learning rate of 0.0001 and a batch size of 32 on a GeForce GTX 1080 Ti graphics processing unit (NVIDIA). For the performance evaluation, 5-fold cross-validation was performed to confirm the generalization ability. The training data set (N=31,940) was randomly shuffled and divided into five equal groups in a stratified manner. Subsequently, four groups were selected to train the model, and the remaining group was used for validation. This process was repeated five times by shifting the internal validation group. Next, we averaged the mean validation costs of the five internal validation groups according to each epoch and found the optimal epoch that provides the lowest validation cost. Then, we retrained the model using the entire training data set with the optimal epoch. The testing data set was evaluated only after the model was completely trained using the training data set. This holdout method provides an unbiased evaluation of the final model by avoiding overfitting to the training data set.

Performance Evaluation and Statistical Evaluation

For each of the different four pretrained models (VGG16, ResNet-50, Inception-v3, and Xception) in FCONet, we evaluated the classification performance based on sensitivity, specificity, and accuracy. More specifically, we calculated true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) based on the groups of COVID-19 pneumonia, other pneumonia, and nonpneumonia. For each group, we expressed measure metrics with the subscripts *covid* for COVID-19, *other* for other pneumonia, and *none* for nonpneumonia, as

where TP_{covid} is the number of COVID-19 testing data correctly classified as COVID-19, TN_{covid} is the number of non–COVID-19 testing data correctly classified as non–COVID-19, FP_{covid} is the number of non–COVID-19 testing data misclassified as COVID-19, and FN_{covid} is the number of COVID testing data misclassified as non–COVID-19. Here, non–COVID-19 testing data include other pneumonia and nonpneumonia. Note that the same calculations were applied to the other pneumonia and nonpneumonia testing data as

We also plotted the receiver operating characteristic (ROC) curve and calculated the area under the curve (AUC) for each of the four different models. Additionally, statistical analysis was performed using MATLAB (R2013b). Analysis of variance (ANOVA) was used to compare differences among COVID-19 pneumonia, non–COVID-19 pneumonia, and nonpneumonia groups. A *P* value less than .001 was considered to indicate statistical significance.

Results

The performance of the FCONet models based on the four pretrained models in the classification of COVID-19 pneumonia, other pneumonia, and nonpneumonia is summarized in Table 4. We compared the metric values of sensitivity (%), specificity (%), and accuracy (%) as well as the AUCs of the four FCONet models based on VGG16, ResNet-50, Inception-v3, and Xception. Based on the testing data, the FCONet models based on ResNet-50, VGG16, and Xception showed excellent classification performance; all these models provided AUC values ranging from 0.99 to 1.00. More specifically, with ResNet-50, the sensitivity, specificity, and accuracy for classifying COVID-19 pneumonia were 99.58%, 100%, and 99.87%, respectively. With VGG16, the sensitivity, specificity, and accuracy for classifying COVID-19 pneumonia were 100%, 99.64%, and 99.75%, respectively. With Xception, the sensitivity, specificity, and accuracy for COVID-19 pneumonia classification were 97.91%, 99.29%, and 98.87%, respectively. For other pneumonia and nonpneumonia, the sensitivity, specificity, and accuracy ranged from 97% to 100% when ResNet-50, VGG16, or Xception was used as the backbone in FCONet. On the other hand, Inception-v3–based FCONet provided relatively low sensitivity, specificity, and accuracy values for all groups of COVID-19 pneumonia, other pneumonia, and nonpneumonia (P<.001).

Table 4. Performance of the FCONet frameworks based on the four pretrained models on the testing data set.

Model and data group	Sensitivity, %	Specificity, %	Accuracy, %	AUC[a]	*P* value
ResNet-50					<.001
COVID-19[b] pneumonia	99.58	100.00	99.87	1.00	
Other pneumonia	97.42	99.81	99.00	0.99	
Nonpneumonia	100.00	98.63	99.12	0.99	
VGG16					<.001
COVID-19 pneumonia	100.00	99.64	99.75	1.00	
Other pneumonia	100.00	99.81	99.87	0.99	
Nonpneumonia	100.00	99.80	99.87	0.99	
Xception					<.001
COVID-19 pneumonia	97.91	99.29	98.87	0.99	
Other pneumonia	98.52	99.05	98.87	0.99	
Nonpneumonia	100.00	100.00	100.00	1.00	
Inception-v3					<.001
COVID-19 pneumonia	88.28	97.68	94.87	0.97	
Other pneumonia	94.10	95.83	95.24	0.98	
Nonpneumonia	98.27	97.25	97.62	0.99	

[a] AUC: area under the curve.

[b] COVID-19: coronavirus disease.

The confusion matrices and ROC curves for the pretrained models on the testing data set are presented in Figures 2-5. More specifically, ResNet-50 exhibited TP_{covid}, TP_{other}, and TP_{none} of 238/239, 268/271, and 289/289, respectively (Figure 2). VGG16 exhibited TP_{covid}, TP_{other}, and TP_{none} of 234/239, 269/271, and 289/289, respectively (Figure 3). Xception exhibited TP_{covid}, TP_{other}, and TP_{none} of 188/239, 257/271, and 289/289, respectively (Figure 4). Inception-v3 exhibited TP_{covid}, TP_{other}, and TP_{none} of 211/239, 255/271, and 284/289, respectively (Figure 5). For the three models of ResNet-50, VGG16, and Xception, the values of AUC were very close to 1 because the predicted probability values were provided as values close to 1 for correct labeling and values close to 0 for incorrect labeling.

On the additional external validation data set, which comprised low-quality CT images of COVID-19 pneumonia embedded in recently published papers, the detection accuracy of ResNet-50 was the highest with 96.97%, followed by Xception (90.71%), Inception-v3 (89.38%), and VGG16 (87.12%) (Table 5).

Figure 2. Confusion matrix and ROC curve in FCONet using ResNet-50; COVID-19: coronavirus disease; ROC: receiver operating characteristic.

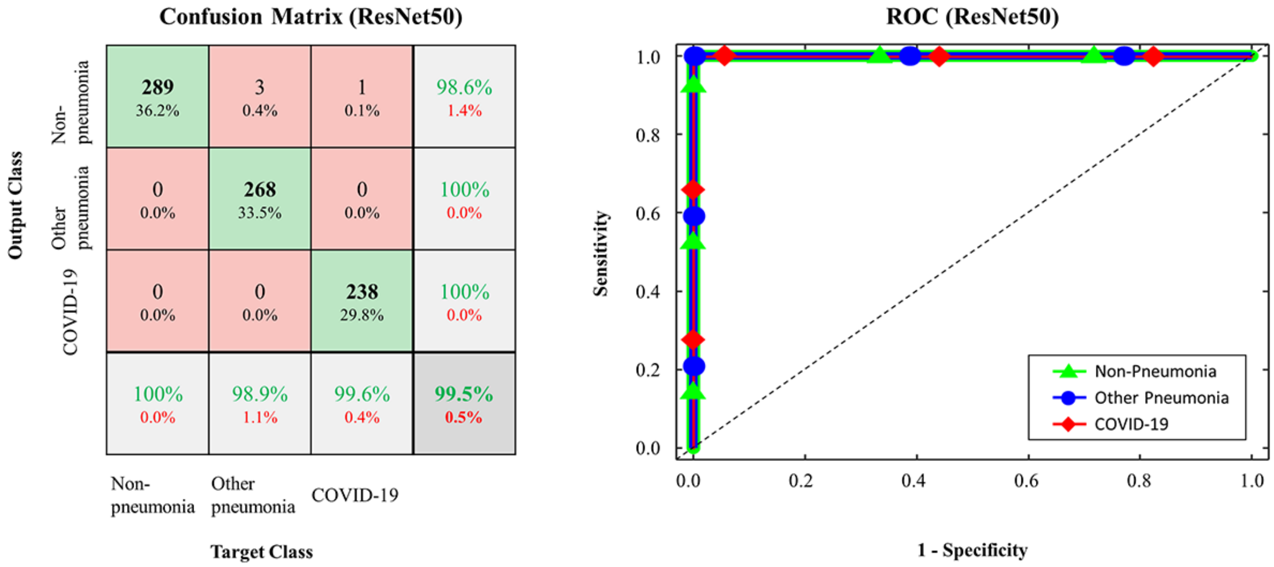
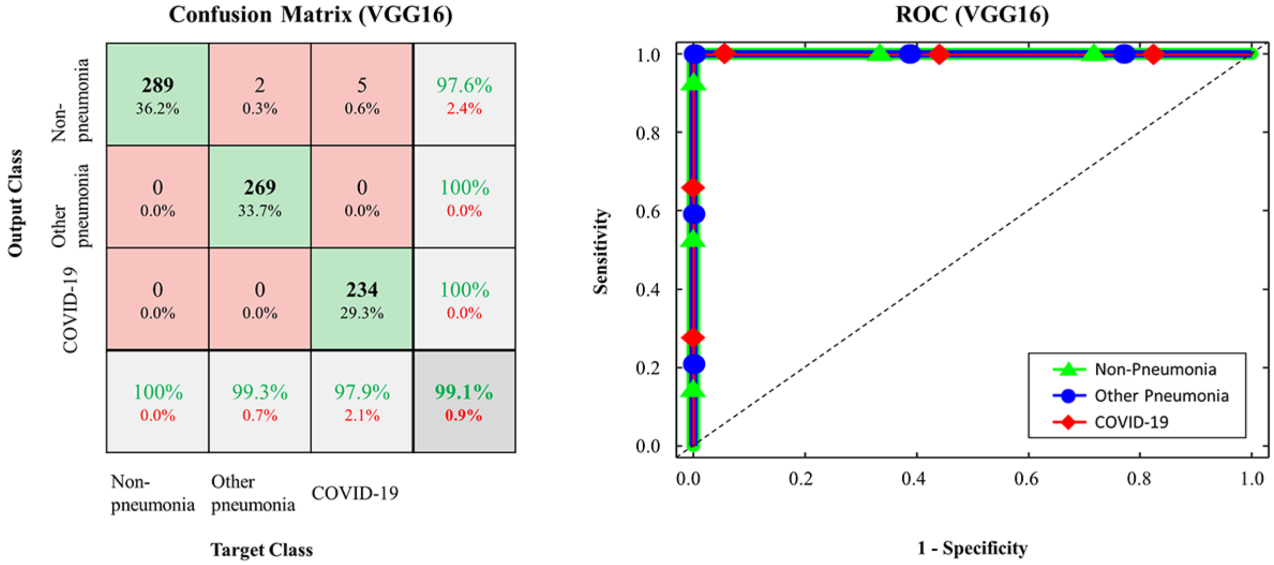
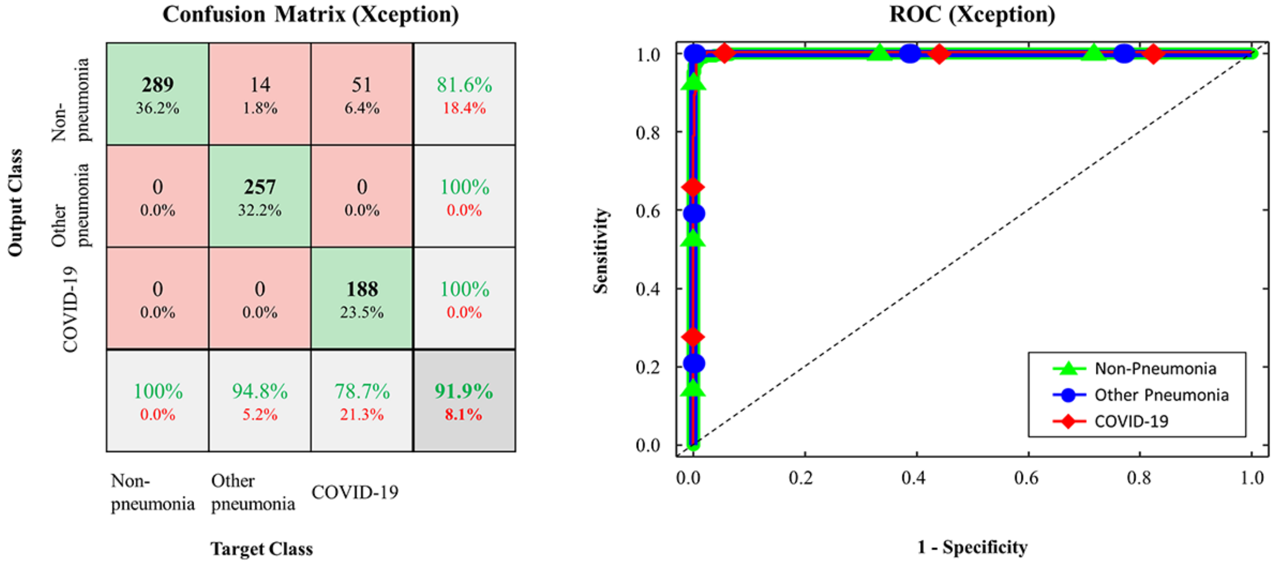


Figure 3. Confusion matrix and ROC curve in FCONet using VGG16; COVID-19: coronavirus disease; ROC: receiver operating characteristic.



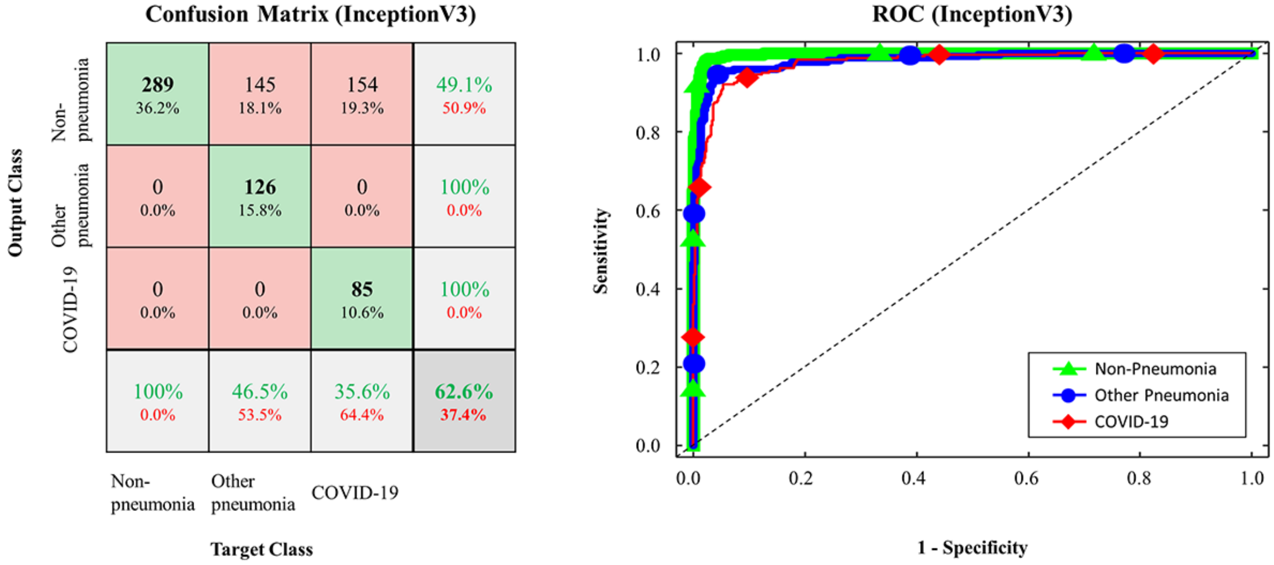
	Non-pneumonia	Other pneumonia	COVID-19	
Non-pneumonia	**289** 36.2%	2 0.3%	5 0.6%	97.6% 2.4%
Other pneumonia	0 0.0%	**269** 33.7%	0 0.0%	100% 0.0%
COVID-19	0 0.0%	0 0.0%	**234** 29.3%	100% 0.0%
	100% 0.0%	99.3% 0.7%	97.9% 2.1%	**99.1%** **0.9%**

Figure 4. Confusion matrix and ROC curve in FCONet using Xception; COVID-19: coronavirus disease; ROC: receiver operating characteristic.



	Non-pneumonia	Other pneumonia	COVID-19	
Non-pneumonia	**289** 36.2%	14 1.8%	51 6.4%	81.6% 18.4%
Other pneumonia	0 0.0%	**257** 32.2%	0 0.0%	100% 0.0%
COVID-19	0 0.0%	0 0.0%	**188** 23.5%	100% 0.0%
	100% 0.0%	94.8% 5.2%	78.7% 21.3%	**91.9%** **8.1%**

Figure 5. Confusion matrix and ROC curve in FCONet using Inception-v3; COVID-19: coronavirus disease; ROC: receiver operating characteristic.



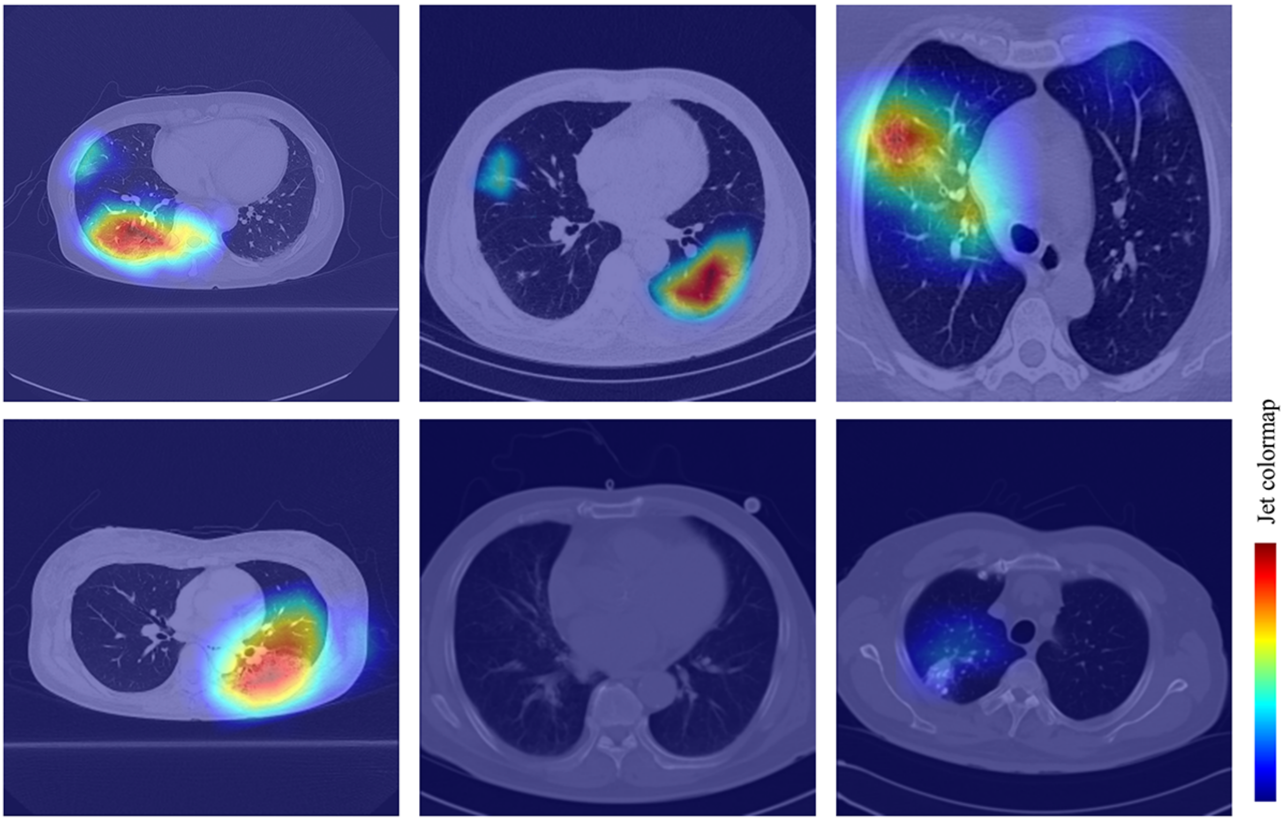
	Non-pneumonia	Other pneumonia	COVID-19	
Non-pneumonia	**289** 36.2%	145 18.1%	154 19.3%	49.1% 50.9%
Other pneumonia	0 0.0%	**126** 15.8%	0 0.0%	100% 0.0%
COVID-19	0 0.0%	0 0.0%	**85** 10.6%	100% 0.0%
	100% 0.0%	46.5% 53.5%	35.6% 64.4%	**62.6%** **37.4%**

Table 5. Performance of each deep learning model on the additional external validation data set of COVID-19 pneumonia images.

Model	Detection accuracy, %
ResNet-50	96.97
VGG16	87.12
Xception	90.71
Inception-v3	89.38

To improve the interpretability of our model, we used the gradient-weighted class activation mapping (Grad-CAM) method [23] to visualize the important regions leading to the decision of FCONet. The model fully generates this localization map without the mapping annotation. The heatmaps (Figure 6) show the suspected regions for the examples of COVID-19, other pneumonia, and nonpneumonia. The heatmaps are standard jet colormaps and are overlapped on the original image, where red color highlights the activation region associated with the predicted class. More specifically, for the COVID-19 image group, the heatmap strongly indicated the suspected regions, as shown in examples from WKUH (Figure 6, top left), CNUH (Figure 6, top middle) and SIRM (Figure 6, top right). For the other pneumonia image groups, the heatmap demonstrated some suspected regions inside the lung area; the intensity was lower than that of the regions in the COVID-19 image group (Figure 6, bottom left). For the healthy image group, there was no heatmap corresponding to the suspected regions (Figure 6, bottom middle). For the lung cancer images, the heatmap indicated some suspected regions inside the lung area; however, the intensity was also lower than that of the regions in the COVID-19 pneumonia group (Figure 6, bottom right).

Figure 6. Confusion matrice and ROC curve in FCONet using VGG16; COVID-19: coronavirus disease; ROC: receiver operating characteristic.



To test the generalizability of our proposed framework, we also trained and tested the models based on institutional data split for COVID-19 data: training data from CNUH and SIRM and tested data from WKUH. Because the COVID-19 data were split with a ratio of 65:35 (773 training data and 421 testing data for COVID-19), the other non–COVID-19 data were randomly split with the same ratio in a stratified fashion. Table 6 summarizes the performance of the FCONet framework. With ResNet-50, the sensitivity, specificity and accuracy for classifying COVID-19 pneumonia were 97.39%, 99.64% and 98.67%, respectively ($P<.001$). With VGG16, the sensitivity, specificity, and accuracy for classifying COVID-19 pneumonia were 97.15%, 99.64% and 98.57%, respectively ($P<.001$). With Xception, the sensitivity, specificity, and accuracy for classifying COVID-19 pneumonia were 90.50%, 94.82% and 92.97%, respectively ($P<.001$). With Inception-v3, the sensitivity, specificity, and accuracy for classifying COVID-19 pneumonia were 74.58%, 99.46% and 88.79%, respectively ($P<.001$). These results show that the FCONet framework can classify COVID-19 regardless of the data split approach.

Table 6. Performance of the FCONet framework based on institutional data split for COVID-19 data.

Model and data group	Sensitivity, %	Specificity, %	Accuracy, %	AUC[a]	*P* value
ResNet-50					<.001
COVID-19 pneumonia	97.39	99.64	98.67	0.99	
Other pneumonia	99.26	98.45	98.637	0.99	
Nonpneumonia	100	100	100	1.0	
VGG16					<.001
COVID-19 pneumonia	97.15	99.64	98.57	0.99	
Other pneumonia	99.26	98.31	98.57	0.99	
Nonpneumonia	100	100	100	1.0	
Xception					<.001
COVID-19 pneumonia	90.50	94.82	92.97	0.98	
Other pneumonia	89.30	94.37	92.97	0.98	
Nonpneumonia	100	100	100	1.0	
Inception-v3					<.001
COVID-19 pneumonia	74.58	99.46	88.79	0.98	
Other pneumonia	97.42	84.93	88.38	0.97	
Nonpneumonia	100	99.42	99.59	0.99	

[a] AUC: area under the curve.

Discussion

Principal Findings

We were able to develop the FCONet deep learning models to diagnose COVID-19 pneumonia in a few weeks using transfer learning based on pretrained models. The FCONet based on ResNet-50 showed excellent diagnostic performance to detect COVID-19 pneumonia. Although the diagnostic accuracy of the FCONet models based on VGG16, ResNet-50, and Xception was excellent in the testing data set (sensitivity, 97.91%, 100%, and 97.91%, respectively; specificity, 100%, 99.64% and 99.29%, respectively), external validation using the low-quality image data set demonstrated that detection accuracy was the highest with ResNet-50 (96.97%), followed by Xception (90.71%), Inception-v3 (89.38%), and VGG16 (87.12%).

To collect as many images as possible within a limited time, we collected readily available chest CT images of COVID-19 patients from institutions in our region (WKUH and CNUH) and a public COVID-19 database established by SIRM. We also systematically searched for chest CT images of COVID-19 embedded in recent papers published between January 19 and March 25, 2020. As these CT images in the published paper were of low quality, we used them only in an additional external validation data set.

During a national crisis such as the COVID-19 pandemic, when the number of infected patients is precipitously increasing and physicians are occupied combating the disease, rapid development of AI methods to detect COVID-19 in CT is crucial to alleviate the clinical burden of physicians and to increase the efficiency of the patient management process [8]. However, significant challenges remain when developing such AI techniques within a limited time to collect CT data and train AI models.

To save time for AI training, we used the chest CT images directly without preprocessing of the lung segmentation. In general, lung segmentation preprocessing is regarded to improve the accuracy of AI training [24-27]; we believe that this improvement can be traded off in exchange for saving time. For AI training, we chose the transfer learning algorithms. Transfer learning enabled us to save time by using pretrained CNN models in the ImageNet data sets, including VGG16, ResNet-50, Inception-v3, and Xception [28]. In our study, FCONet based on ResNet-50 showed excellent results and outperformed the FCONet models based on the other three pretrained models in both our testing data set and the additional external validation data set. The VGG model is regarded as a traditional sequential network architecture and may be hampered by slow training and a large model size [17]. The ResNet-50 model is characterized by network-in-network architectures, which have much deeper layers than those of VGG models, enabling reduction of the model size [18]. Our results suggest that transfer learning for a 2D deep learning framework can be robustly applied to deep learning models and that the ResNet-50 model provides the best accuracy.

We adopted AI training based on a 2D image framework rather than a 3D framework because 3D deep learning requires significantly higher computation power than sequential 2D image analyses [29]. In our emergent clinical setting to fight COVID-19, a simple and rapid model may be preferable to a complex and slow model. In addition, training a 2D image framework saves time for AI development.

Despite limited resources and time, we were able to generate a deep learning model to detect COVID-19 from chest CT with excellent diagnostic accuracy. To date, a few papers have been published on AI models for detecting COVID-19 in chest CT images [6]. An AI model named COVNet was trained using 4356 CT images from six hospitals in China. It showed 90% sensitivity (95% CI 83%-94%) and 96% specificity (95% CI 93%-98%) in detecting COVID-19, which is comparable with our results. However, we cannot compare our FCONet to COVNet because the training and testing data sets are different.

Although chest radiography is the most commonly used imaging tool to detect COVID-19, its sensitivity is lower than that of CT [30]. However, in this pandemic period, clinicians may hesitate to perform chest CT due to limited resources such as CT scanners and radiologists as well as contamination of CT scanners [31]. In our hospitals (WKUH and CNUH), we recently dedicated a mobile CT scanner exclusively to COVID-19 patients to alleviate the physical and mental stress of medical staff. We believe that incorporating an AI model to detect suspicious lesions of COVID-19 pneumonia can improve the workflow by providing rapid diagnostic support.

Limitations and Future Work

Our study has several limitations. Firstly, our AI models were validated mainly using a split testing data set. Thus, the testing data set was obtained from the same sources as the training data set. This may raise issues of generalizability and overfitting of our models [32,33]. Indeed, the detection accuracy of our model decreased slightly for the external validation data set using chest CT images from published papers. However, the initial goal was to incorporate a deep learning model in our emergent clinical setting as a supporting tool. In the near future, we will train our model using CT images from various institutions and countries. Secondly, we used a relatively small amount of data to train the deep learning models. Thus, we will establish a sustainable AI training system that can continue to train our model using prospectively collected CT images.

Conclusions

We described FCONet, a simple 2D deep learning framework based on a single chest CT image, as a diagnostic aid that provides excellent diagnostic performance to diagnose COVID-19 pneumonia. The FCONet model based on ResNet-50 appears to be the best model, outperforming other models based on VGG16, Xception, and Inception-v3.

Acknowledgments

This work was partially supported by the National Research Foundation of Korea, Ministry of Science, ICT, and Future Planning, through the Basic Science Research Program (grant number: NRF-2020R1A2C1014829), and from the Korea Health Industry Development Institute (grant number: HI18C2383 and HI18C1216).

Conflicts of Interest

None declared.

References

1. Coronavirus disease 2019 (COVID-19) Situation Report – 84. World Health Organization. 2020 Apr 13. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200413-sitrep-84-covid-19.pdf?sfvrsn=44f511ab_2 [accessed 2020-06-23]
2. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, et al. Covid-19 in Critically Ill Patients in the Seattle Region - Case Series. N Engl J Med 2020 May 21;382(21):2012-2022 [FREE Full text] [doi: 10.1056/NEJMoa2004500] [Medline: 32227758]
3. Loeffelholz MJ, Tang Y. Laboratory diagnosis of emerging human coronavirus infections - the state of the art. Emerg Microbes Infect 2020 Dec;9(1):747-756 [FREE Full text] [doi: 10.1080/22221751.2020.1745095] [Medline: 32196430]
4. Li Y, Yao L, Li J, Chen L, Song Y, Cai Z, et al. Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. J Med Virol 2020 Jul;92(7):903-908 [FREE Full text] [doi: 10.1002/jmv.25786] [Medline: 32219885]
5. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020 Feb 26:200642 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
6. Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology 2020 Jul;296(1):172-180 [FREE Full text] [doi: 10.1148/radiol.2020201365] [Medline: 32255413]
7. Vashist SK. In Vitro Diagnostic Assays for COVID-19: Recent Advances and Emerging Trends. Diagnostics (Basel) 2020 Apr 05;10(4):202 [FREE Full text] [doi: 10.3390/diagnostics10040202] [Medline: 32260471]
8. Li D, Wang D, Dong J, Wang N, Huang H, Xu H, et al. False-Negative Results of Real-Time Reverse-Transcriptase Polymerase Chain Reaction for Severe Acute Respiratory Syndrome Coronavirus 2: Role of Deep-Learning-Based CT Diagnosis and Insights from Two Cases. Korean J Radiol 2020 Apr;21(4):505-508 [FREE Full text] [doi: 10.3348/kjr.2020.0146] [Medline: 32174053]

9. Long C, Xu H, Shen Q, Zhang X, Fan B, Wang C, et al. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? Eur J Radiol 2020 May;126:108961 [FREE Full text] [doi: 10.1016/j.ejrad.2020.108961] [Medline: 32229322]
10. Mossa-Basha M, Medverd J, Linnau K, Lynch JB, Wener MH, Kicska G, et al. Policies and Guidelines for COVID-19 Preparedness: Experiences from the University of Washington. Radiology 2020 Apr 08:201326 [FREE Full text] [doi: 10.1148/radiol.2020201326] [Medline: 32267209]
11. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. Radiology 2020 Mar 10:200823 [FREE Full text] [doi: 10.1148/radiol.2020200823] [Medline: 32155105]
12. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology 2020 Mar 19:200905 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
13. COVID-19 Database. Italian Society of Medical and Interventional Radiology (SIRM). URL: https://www.sirm.org/en/category/articles/covid-19-database/ [accessed 2020-06-23]
14. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-dataset: a CT scan dataset about COVID-19. arXiv. 2020 Mar 20. URL: https://arxiv.org/abs/2003.13865 [accessed 2020-06-23]
15. Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. Eur Radiol 2002 May;12(5):1052-1057. [doi: 10.1007/s003300101126] [Medline: 11976846]
16. Ohana M, Labani A, Severac F, Jeung M, Gaertner S, Caspar T, et al. Single source dual energy CT: What is the optimal monochromatic energy level for the analysis of the lung parenchyma? Eur J Radiol 2017 Mar;88:163-170. [doi: 10.1016/j.ejrad.2017.01.016] [Medline: 28189203]
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. 2014. URL: https://arxiv.org/abs/1409.1556 [accessed 2020-06-23]
18. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 26-July 1, 2016; Las Vegas, NV. [doi: 10.1109/CVPR.2016.90]
19. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 26-July 1, 2016; Las Vegas, NV. [doi: 10.1109/cvpr.2016.308]
20. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 Presented at: 017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI. [doi: 10.1109/cvpr.2017.195]
21. Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput 2017 Sep;29(9):2352-2449. [doi: 10.1162/neco_a_00990]
22. Kingma DP, Ba JL. Adam: A method for stochastic optimization. 2015 Presented at: 3rd International Conference for Learning Representations; May 7-9, 2015; San Diego, CA URL: https://arxiv.org/pdf/1412.6980.pdf
23. Selvaraju R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. : Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision; 2017 Presented at: 2017 IEEE International Conference on Computer Vision (ICCV); Oct. 22-29, 2017; Venice, Italy. [doi: 10.1109/iccv.2017.74]
24. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology 2020 Mar 19:200905 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
25. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv. URL: https://arxiv.org/abs/2002.09334 [accessed 2020-06-23]
26. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label. medRxiv 2020 Mar 26:preprint. [doi: 10.1101/2020.03.12.20027185]
27. Chung H, Ko H, Jeon SJ, Yoon K, Lee J. Automatic Lung Segmentation With Juxta-Pleural Nodule Identification Using Active Contour Model and Bayesian Approach. IEEE J Transl Eng Health Med 2018;6:1-13. [doi: 10.1109/jtehm.2018.2837901]
28. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. Imagenet: A large-scale hierarchical image database; 2009 Presented at: 2009 IEEE Conference on Computer Vision and Pattern Recognition; June 20-25, 2009; Miami, FL p. A. [doi: 10.1109/cvpr.2009.5206848]
29. Maturana D, Scherer S. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. 2015 Presented at: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 28 September-3 October 2015; Hamburg, Germany. [doi: 10.1109/iros.2015.7353481]
30. Yoon SH, Lee KH, Kim JY, Lee YK, Ko H, Kim KH, et al. Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in Korea. Korean J Radiol 2020 Apr;21(4):494-500 [FREE Full text] [doi: 10.3348/kjr.2020.0132] [Medline: 32100485]

31. Mossa-Basha M, Meltzer CC, Kim DC, Tuite MJ, Kolli KP, Tan BS. Radiology Department Preparedness for COVID-19: Scientific Expert Panel. Radiology 2020 Mar 16:200988 [FREE Full text] [doi: 10.1148/radiol.2020200988] [Medline: 32175814]
32. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. Korean J Radiol 2019 Mar;20(3):405-410 [FREE Full text] [doi: 10.3348/kjr.2019.0025] [Medline: 30799571]
33. Santosh KC. AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. J Med Syst 2020 Mar 18;44(5):93 [FREE Full text] [doi: 10.1007/s10916-020-01562-1] [Medline: 32189081]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
API: application programming interface
AUC: area under the curve
CNN: convolutional neural networks
CNUH: Chonnam National University Hospital
COVID-19: coronavirus disease
CT: computed tomography
DICOM: Digital Imaging and Communications in Medicine
FCONet: fast-track COVID-19 classification network
FN: false negative
FP: false positive
Grad-CAM: gradient-weighted class activation mapping
ROC: receiver operating characteristic
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SIRM: Italian Society of Medical and Interventional Radiology
TN: true negative
TP: true positive
WKUH: Wonkwang University Hospital

Edited by G Eysenbach; submitted 23.04.20; peer-reviewed by J Lee, T Goodwin; comments to author 14.05.20; revised version received 31.05.20; accepted 21.06.20; published 29.06.20.

Please cite as:
Ko H, Chung H, Kang WS, Kim KW, Shin Y, Kang SJ, Lee JH, Kim YJ, Kim NY, Jung H, Lee J
COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep Learning Framework With a Single Chest CT Image: Model Development and Validation
J Med Internet Res 2020;22(6):e19569
URL: http://www.jmir.org/2020/6/e19569/
doi:10.2196/19569
PMID:32568730

Original Paper

A COVID-19 Risk Assessment Decision Support System for General Practitioners: Design and Development Study

Ying Liu[1*], MD; Zhixiao Wang[2*], BSc; Jingjing Ren[1*], MD, PhD; Yu Tian[2], PhD; Min Zhou[1], ME; Tianshu Zhou[2], PhD; Kangli Ye[1], MBBS; Yinghao Zhao[3], MSc; Yunqing Qiu[1], MD, PhD; Jingsong Li[2,3], PhD

[1]The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China

[2]Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

[3]Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Jingsong Li, PhD
Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education
College of Biomedical Engineering and Instrument Science
Zhejiang University
38 Zheda Road
Hangzhou, 310027
China
Phone: 86 57187951564
Email: ljs@zju.edu.cn

Abstract

Background: The coronavirus disease (COVID-19) has become an urgent and serious global public health crisis. Community engagement is the first line of defense in the fight against infectious diseases, and general practitioners (GPs) play an important role in it. GPs are facing unique challenges from disasters and pandemics in delivering health care. However, there is still no suitable mobile management system that can help GPs collect data, dynamically assess risks, and effectively triage or follow-up with patients with COVID-19.

Objective: The aim of this study is to design, develop, and deploy a mobile-based decision support system for COVID-19 (DDC19) to assist GPs in collecting data, assessing risk, triaging, managing, and following up with patients during the COVID-19 outbreak.

Methods: Based on the actual scenarios and the process of patients using health care, we analyzed the key issues that need to be solved and designed the main business flowchart of DDC19. We then constructed a COVID-19 dynamic risk stratification model with high recall and clinical interpretability, which was based on a multiclass logistic regression algorithm. Finally, through a 10-fold cross-validation to quantitatively evaluate the risk stratification ability of the model, a total of 2243 clinical data consisting of 36 dimension clinical features from fever clinics were used for training and evaluation of the model.

Results: DDC19 is composed of three parts: mobile terminal apps for the patient-end and GP-end, and the database system. All mobile terminal devices were wirelessly connected to the back end data center to implement request sending and data transmission. We used low risk, moderate risk, and high risk as labels, and adopted a 10-fold cross-validation method to evaluate and test the COVID-19 dynamic risk stratification model in different scenarios (different dimensions of personal clinical data accessible at an earlier stage). The data set dimensions were (2243, 15) when only using the data of patients' demographic information, clinical symptoms, and contact history; (2243, 35) when the results of blood tests were added; and (2243, 36) after obtaining the computed tomography imaging results of the patient. The average value of the three classification results of the macro–area under the curve were all above 0.71 in each scenario.

Conclusions: DCC19 is a mobile decision support system designed and developed to assist GPs in providing dynamic risk assessments for patients with suspected COVID-19 during the outbreak, and the model had a good ability to predict risk levels in any scenario it covered.

(J Med Internet Res 2020;22(6):e19786) doi:10.2196/19786

KEYWORDS

COVID-19; dynamic risk stratification; decision support system; telemedicine triage system; multiclass logistic regression

Introduction

Background

In December 2019, cases of pneumonia with unknown cause, which was designated the coronavirus disease (COVID-19) in February by the World Health Organization (WHO), were reported in Wuhan, Hubei Province, China [1]. The number of cases of COVID-19 increased rapidly worldwide over the next few months, and the WHO announced a pandemic on March 11, 2020 [2]. As of April 29, 2020, 2,995,758 cases of infection and 204,987 deaths were reported worldwide [3]. COVID-19 has become an urgent and serious global public health crisis.

As a new infectious disease in humans caused by severe acute respiratory syndrome coronavirus 2 (SARS-COV-2), COVID-19 is characterized by respiratory symptoms and human-to-human transmission [4]. Rapid asymptomatic transmission of COVID-19 and high mortality in susceptible, elderly, and immunocompromised people make it necessary to identify and control positive patients quickly [5]. However, some characteristics of SARS-COV-2 have further increased the difficulty of early identification and isolation of patients. Wu et al's [6] research showed that the mean basic reproduction number of SARS-CoV-2 is 2.68 (95% credible interval [CrI] 2.47-2.86) and that the epidemic doubling time is 6.4 days (95% CrI 5.8-7.1 days). Viral RNAs can be found in nasal discharge, sputum, and sometimes in blood or feces [7]. As described in past studies, similar to the other respiratory viral infections, the most common symptoms at the onset of COVID-19 were fever, cough, and fatigue [4,7], which increased difficulties in early identification of patients who were infected. Meanwhile, with the development of the epidemic, people's panic and the shortage of medical resources has greatly hindered the prevention and control of the disease. In addition, research by Ji et al [8] revealed that the current mortality of COVID-19 is significantly positively related to the burden of health care. Therefore, how to evaluate patients with COVID-19, suspected cases, and other patients with similar symptoms, and effectively triage under the condition of medical resource shortages is of great significance to protect the susceptible population, reduce hospital cross infection, and decrease the burden on medical resources.

Challenges to General Practitioners in the COVID-19 Outbreak

Community engagement is the first line of defense for effective prevention and control of infectious diseases [9]. In addition, in China, general practitioners (GPs) in primary care institutions play an important role in this battle. GPs are engaged in blocking the viral transmission by monitoring people at designated checkpoints, treating patients, and providing medical surveillance in the community. In addition, they detect, diagnose, and treat patients at different levels of fever clinics, providing continuous care for patients who are discharged and are chronically ill [9].

Disasters and pandemics pose unique challenges to health care delivery [10]. There are a lot of challenges for GPs and other health professionals to maintain their work and meet the increasing need for digital health care. First, the most serious problem is that they are at risk of getting infected; therefore, in-person visits need to be reduced to a minimum level. Second, many health service centers have experienced a rush of people who are worried or infected, and patients with chronic diseases who are undergoing treatment or observation are also at risk of being infected when they visit the outpatient clinics. Finally, medical information systems are generally interrupted and discontinuous; data still needs to be manually recorded or repeatedly registered by health professionals, which increases the workload of health professionals and makes it impossible to manage the dynamic information of patients in a timely and comprehensive manner.

The Role of Telemedicine in the Era of COVID-19

Telemedicine has a critical role in emergency responses and is an ideal model for managing infectious diseases [10,11]. In addition, the key factor that slows the spread of the virus is the "social distance," which can directly reduce person-to-person contact [11]. Telemedicine can also deploy large numbers of health care providers rapidly to facilitate triaging without in-person visits, provide clinical services when local medical resources are unable to meet the demand, and reduce the risk of nosocomial infection.

Until now, researchers have developed many different forms of telemedicine systems to meet the needs of fighting against the epidemic. Jin et al [12] built and deployed an artificial intelligence (AI)–assisted system for automatic computed tomography (CT) image analysis and recognition of COVID-19 within 4 weeks, which has been applied in many hospitals, greatly reducing the pressure of radiologists. Du et al [13] proposed a hybrid AI model for COVID-19 prediction that fully considered the effects of prevention and control measures, and the improvement of public prevention awareness, and has been applied in a number of Chinese cities. Moreover, Reeves et al [14] built a series of standardized tools based on the existing electronic health record (EHR) system for the COVID-19 epidemic to support outbreak management, including scripted triaging, electronic check-in, standard ordering and documentation, secure messaging, real time data analytics, and telemedicine capabilities.

These models and systems have considered the main problems from different perspectives during the COVID-19 epidemic and have been deployed and published in a short time, which has played a significant role in fighting the epidemic. However, to the best of our knowledge, there is still no suitable mobile management system that can help GPs realize the automatic collection of data, dynamic risk assessment, and effective triaging and follow-ups with patients with COVID-19, as well as the effective reduction of the pressure on large designated general hospitals.

Study Aims

In this study, by integrating doctor experience, clinical guidelines, and retrospective data, we designed and developed a dynamic risk assessment decision support system for COVID-19 (DDC19) to assist GPs in data collection, dynamic risk assessment, triage management, and follow-ups during the outbreak of COVID-19. The DDC19 is designed to build a free mobile app that can cover all the different situations encountered by residents and GPs, and GPs can use it for dynamic continuity management. We describe our experiences, lessons learned, and recommendations for the design and implementation of telemedicine tools in future health emergencies.

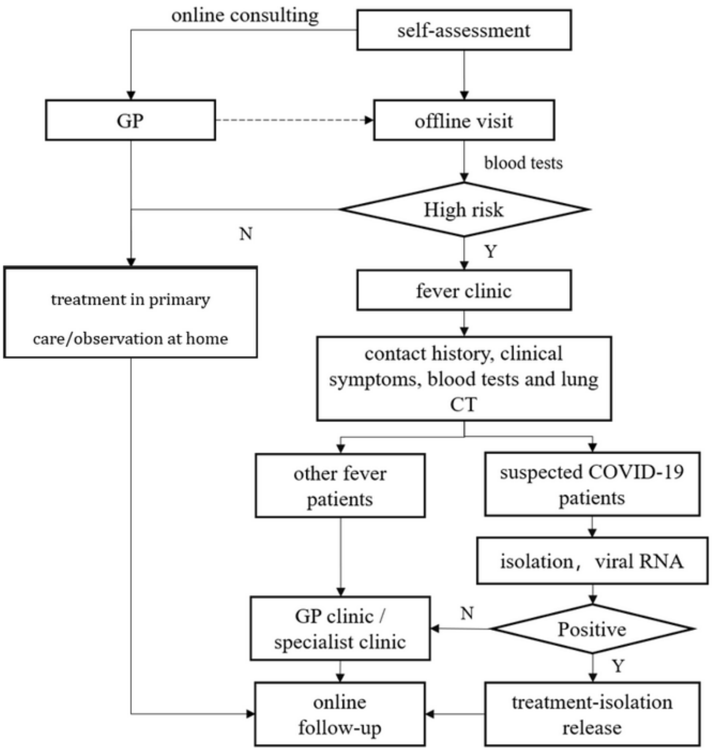
Methods

The Covering Scenes of DDC19

To achieve the early assessment and triage of patients with COVID-19 and ease the pressure of shortages in medical resources, the following key issues need to be resolved: how to fully grasp and effectively manage the residents' status in real time without increasing the GP's working burden and, without omitting potential patients with COVID-19, how to effectively use medical knowledge and risk stratification models to achieve effective evaluation and classification, as well as the patients' scientific stratification.

Accordingly, based on the principle of the four early approaches (early detection, early reporting, early isolation, and early treatment) and the actual scenes and process of patients using health care, DDC19 was designed to help GPs manage their patients who had a fever or respiratory symptoms, or suspected infection with SARS-CoV-2 during the outbreak of COVID-19. Several scenarios were involved and are demonstrated in Figure 1. First, patients should fill in the self-assessment questionnaire in the system. After that, according to the risk level assessed by the system and the diagnosis and treatment suggestions given, patients can choose to contact the GPs online for consultations or directly go to the primary care center for further examination and treatment. Second, when patients choose to visit in-person, those who are at low- and moderate-risk may receive blood tests and have their risks reassessed based on actual clinical symptoms and blood test results. After reassessment, if patients are still at low- and moderate-risk, GPs will give them some necessary treatment and follow-up online. For patients who are high risk, they should be transferred to a fever clinic as soon as possible. Third, patients at high risk who are transferred to the fever clinic will receive a lung CT scan and viral RNA test. If the lung CT shows a clinical manifestation of COVID-19 or the viral RNA test is positive, they will be isolated and treated. Otherwise, they will be transferred to the GP or specialist that is needed. Finally, the cured patients will return to the community, GPs will continue monitoring their conditions and dynamically assess risk online, as well as follow-up with them offline if needed.

Figure 1. The covering scenes of patients using health care during the outbreak of COVID-19. COVID-19: coronavirus disease; CT: computed tomography; GP: general practitioner.

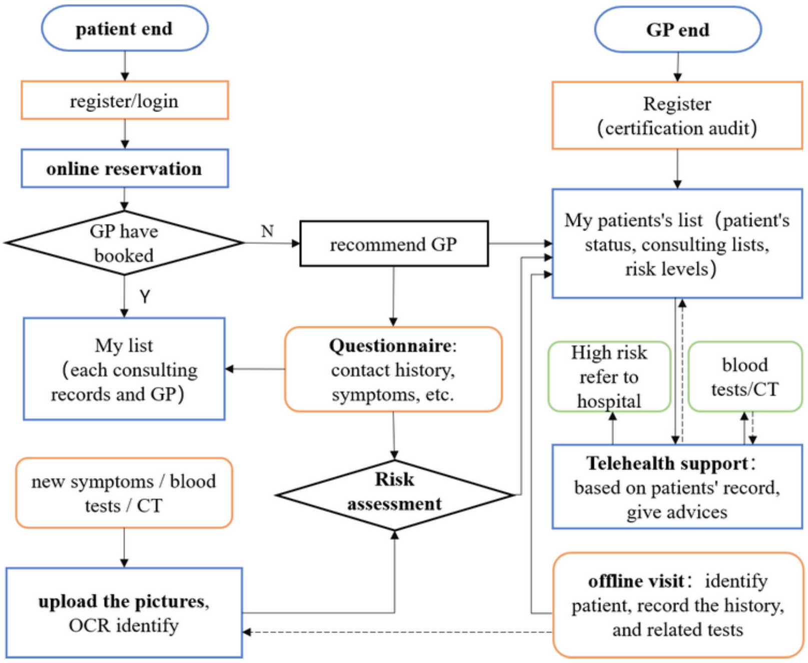


The Design of Main Functions and Business Process of DDC19

To make full use of the previously mentioned online-offline combined triage mode during the COVID-19 outbreak, we deeply integrated GP's experience and suggestions, and designed the business process of the DDC19 system in detail. In the first step, we reviewed the COVID-19 guidelines [15] and fully integrated the actual clinical experience of GPs on the frontline of COVID-19 prevention and control to determine the main functions of the system and the content of the questionnaire. In the second step, based on the content of the questionnaire, the main data elements of the risk stratification model were determined, and, further considering the variability of the amount of the available questionnaire content data, a dynamic risk stratification model was constructed. In the third step, we validated and adjusted the model with retrospective clinical data. Therefore, the functions and business processes of the system were gradually completed according to the changes of clinical requirements.

As shown in Figure 2, the system mainly includes the patient-end and the doctor-end. Its main functions, including online patient clinical data collection, dynamic risk assessment, online classification, and appointment, can be achieved through the interaction between two mobile terminals. For the online collection module of the patient's clinical data, it was completed by the structured questionnaires and uploading of examination report pictures. The questionnaire content related to the early assessment of COVID-19 will be gradually adjusted according to the experience of GPs and clinical guidelines [15]. The detailed information of the health questionnaire is shown in Multimedia Appendix 1. To ensure that patients with COVID-19 were not omitted and that enough attention was given to the patients with low risk and moderate risk, and to improve the reliability of risk assessment, a dynamic risk assessment module was designed based on the current clinical practice experience and expert recommendations, and the risk assessment level of COVID-19 was divided into low risk, moderate risk, and high risk. In addition, to ensure the patients' risk level can be provided in real time and to consider the different dimensions of clinical data that patients uploaded, the dynamic risk stratification model, based on machine learning, was trained by the retrospective data (see details in the next section).

Figure 2. The main business flowchart of DDC19. CT: computed tomography; GP: general practitioner; OCR: optical character recognition.



The Construction of the COVID-19 Dynamic Risk Stratification Model

Considering the characteristics of COVID-19 such as unobvious symptom specificity and long incubation period, we constructed a COVID-19 dynamic risk stratification model with high recall and clinical interpretability. Based on a multiclass logistic regression algorithm, it integrates the retrospective clinical data analysis results of patients, doctors' experiences, and clinical guidelines.

To meet the risk assessment requirements of COVID-19 at the system level for multiple levels and multiple scenarios, this model was based on patients' data from the fever clinic during the COVID-19 epidemic period and patients' risk levels provided by GPs according to COVID-19 diagnosis and treatment guidelines [15]. Currently, for patients who visit the fever clinic of designated hospitals during the epidemic, GPs will provide the COVID-19 risk assessment to them based on epidemiological history and clinical manifestations (symptoms, blood tests, and lung CT scan). The detailed content and process of evaluation are shown in Textbox 1. According to the different levels of health information submitted by patients in the system, the research team used the patient's risk level (assessed by GPs according to clinical guidelines as patient's label), demographic data, epidemiological history, clinical symptoms, data of laboratory tests, and lung CT imaging as the characteristics of the patient to construct a dynamic risk stratification model. The detailed model construction process is shown in Figure 3.

Compared with the fixed-risk assessment method, which required all the information of patient's medical history, laboratory examination, and the result of the lung CT, this model can not only provide COVID-19 risk stratification for new patients but also provide a dynamic risk level assessment and medical advice based on the clinical data submitted in the system by the patient at the different stages.

Textbox 1. The content and standards of coronavirus disease risk assessment.

Epidemiological history

- Within 14 days before the onset of the disease, the patient has a travel or residence history in the high-risk regions or countries.
- Within 14 days before the onset of the disease, the patient has a history of contact with those infected with severe acute respiratory syndrome coronavirus 2 (those with a positive nucleic acid testing result).
- Within 14 days before the onset of the disease, the patient had direct contact with patients with fever or respiratory symptoms in high-risk regions or countries.
- Disease clustering (2 or more cases with fever or respiratory symptoms occur at such places as homes, offices, and school classrooms, within 2 weeks)

Clinical manifestations

- Fever or respiratory symptoms
- The white blood cell counts in the early stage of the disease is normal or decreased, or the lymphocyte count decreases over time.
- Computed tomography imaging features of the coronavirus disease

Figure 3. The construction process of the dynamic risk stratification model.

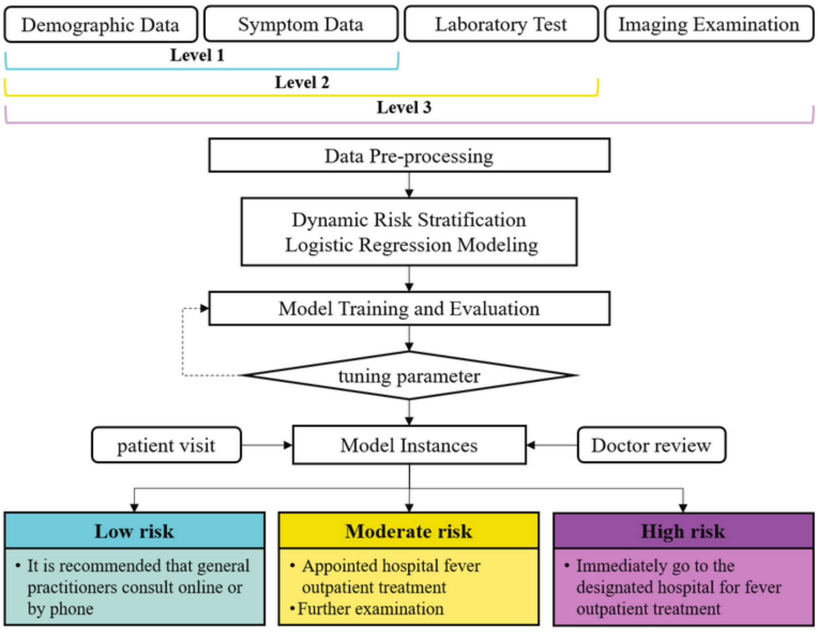


Table 1 shows the data elements used in the dynamic risk stratification model and their corresponding detailed descriptions. Data elements mainly include four categories: demographic data, epidemiological history, clinical symptoms (eg, fever, cough), the data of laboratory tests (eg, blood routine test and C-reactive protein [CRP] test), and lung CT imaging data. According to several studies [16-19], these data elements include the early clinical symptoms of COVID-19 and the easily accessible data of clinical examination, which plays an important role in early assessment.

Table 1. The data elements in the dynamic risk stratification model.

Main category, data element	Description
Patient demographic information	
Patient ID	Unique patient identifier
Gender	Patient’s gender identity
Age	Patient’s age
Clinical symptoms	
Fever	Normal: temprature≤37.2˚C; low fever: temperature between 37.2˚C and 38.5˚C; High fever: temperature≥38.5˚C
Cough	Cough or dry cough
Sputum production	N/A[a]
Fatigue	N/A
Breathing	Shortness of breath, anhelation, polypnea, etc
Chest uncomfortable	Chest pain or chest distress
Pharyngalgia	Pharyngalgia
Headache	Headache or dizziness
Chills	Fear of cold
Soreness	Body aches, joint pain, myalgia
Stuffy nose	Stuffy nose or runny nose
Gastrointestinal reactions	Feeling sick, vomiting, abdominal pain, diarrhea, etc
Epidemiology history	
Contact history	Have a COVID-19[b] contact history
Imaging examination	
CT[c]	Lung CT shows viral pneumonia
Blood routine examination	
WBC[d]	White blood cell count (10E9/L)
GRAN[e]	Neutrophil count (10E9/L)
LYM[f]	Lymphocyte count (10E9/L)
RBC[g]	Red blood cell count (10E12/L)
HGB[h]	Hemoglobin concentration (g/L)
HCT[i]	Hematocrit (%)
MCV[j]	Mean corpuscular volume (fl)
MCH[k]	Mean hemoglobin content (pg)
MCHC[l]	Mean corpuscular hemoglobin concentration (g/L)
RDW[m]	Red blood cell distribution width (%)
PLT[n]	Blood platelet count (10E9/L)
MPV[o]	Mean platelet volume (fl)
PCT[p]	Platelet hematocrit (%)
PDW[q]	Platelet distribution width (10 [GSD[r]])
MO[s]	Mononuclear cell count (10E9/L)
EO[t]	Eosinophil count (10E9/L)

Main category, data element	Description
BA[u]	Basophil count (10E9/L)
NRBC[v]	Percentage of nucleated red blood cells
IG[w]	Immature granulocyte percentage (%)
CRPH[x]	C-reactive protein (mg/L)

[a] Not applicable.
[b] COVID-19: coronavirus disease.
[c] CT: computed tomography.
[d] WBC: white blood cell.
[e] GRAN: granulocytes.
[f] LYM: lymphocyte.
[g] RBC: red blood cell.
[h] HGB: hemoglobin.
[i] HCT: hematocrit.
[j] MCV: mean corpuscular volume.
[k] MHC: mean hemoglobin content.
[l] MCHC: mean corpuscular hemoglobin concentration.
[m] RDW: red blood cell distribution width.
[n] PLT: platelet.
[o] MPV: mean platelet volume.
[p] PCT: platelet hematocrit.
[q] PDW: platelet distribution width.
[r] GSD: geometric standard deviation.
[s] MO: mononuclear.
[t] EO: eosinophil.
[u] BA: basophil.
[v] NRBC: nucleated red blood cells.
[w] IG: immature granulocyte.
[x] CRPH: C-reactive protein.

After the optical character recognition (OCR) module and natural language processing (NLP) module recognize and preprocess the image and text data, the structured patient data is extracted from the app back end database to the data preprocessing module for subsequent data analysis; the detailed data elements are shown in Table 1. The system established a multiclass logistic regression model to dynamically predict and stratify the risk of COVID-19 infection based on the input of different levels of patient clinical data. Logistic regression has good classification ability and interpretability, which has been widely used in the field of machine learning. The hypothetical function is:

In this formula, x is the model input, and θ is the main parameter of the model. This parameter is obtained mainly by fitting the observed data features and corresponding labels. For the case of linear boundaries, $\theta^T x$ can be explained as follows:

The output $h_\theta(X) \in (0,1)$ represents the probability that the result is 1. Therefore, for the binary classification problem, the probability that the result of the input $x$classification is category 1 and category 0 is:

$$P(y = 1 \mid x; \theta) = h_\theta(x)$$

$$P(y = 0 \mid x; \theta) = 1 - h\theta(x)$$

To measure the difference between the model prediction classification result $h_\theta(X)$ and the true value y, the corresponding loss function is defined as:

For m training samples, the loss function is expressed as follows:

Based on the retrospective data of m patients, by minimizing the loss function $J(\theta)$, the parameter θ makes the predicted classification category closest to the patient's true risk level obtainable. However, to balance the distribution of medical resources as much as possible and allow more patients to get enough medical attention, the risk stratification is set to three categories: low risk, moderate risk, and high risk. Therefore, in terms of model construction, to adapt to the above multiclassification situation, the training set data is divided into two different categories to train multiple classifiers. Assuming that n represents the number of categories in the training set, then classifiers need to be trained for prediction; the function is expressed as:

Where i is the i-th category. For the input data of the newly visited patient, classifiers will be used to predict the risk level, and the classification result with the highest output probability is taken as the risk level of the patient. The function is expressed as:

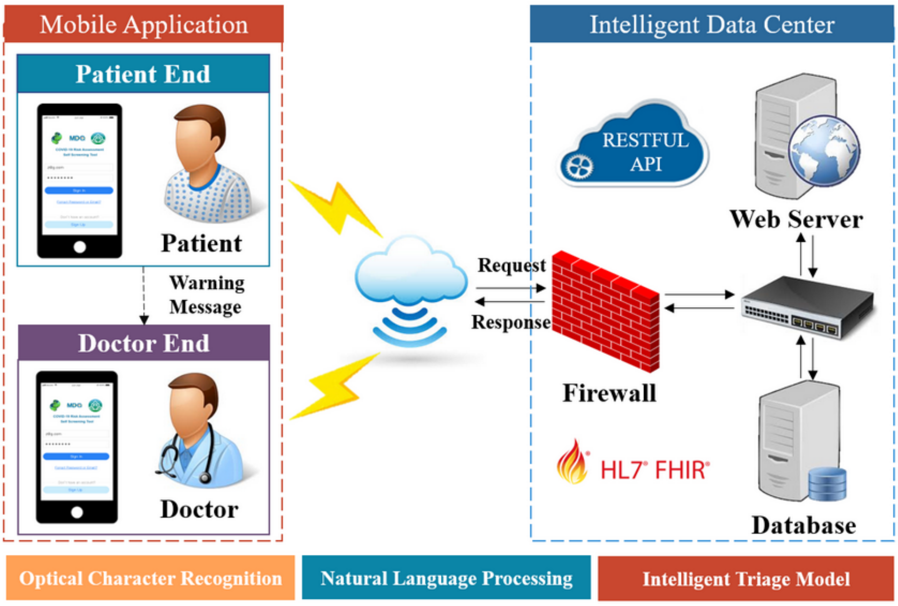
Results

The Architecture and Prototype of DDC19

To prove the effectiveness of the previously mentioned COVID-19 dynamic risk stratification model, as well as the feasibility and practicability of the early assessment and triage system, the research team developed the system prototype and put it into practical operation and application.

As shown in Figure 4, the DDC19 is mainly composed of three parts: two main mobile terminal apps (patient-end and GP-end), and the database system with its related components and underlying related support model. All mobile terminal devices connect to the back end data center wirelessly to achieve request sending and data transmission. In this section, the main architecture and functions of each part will be described in detail.

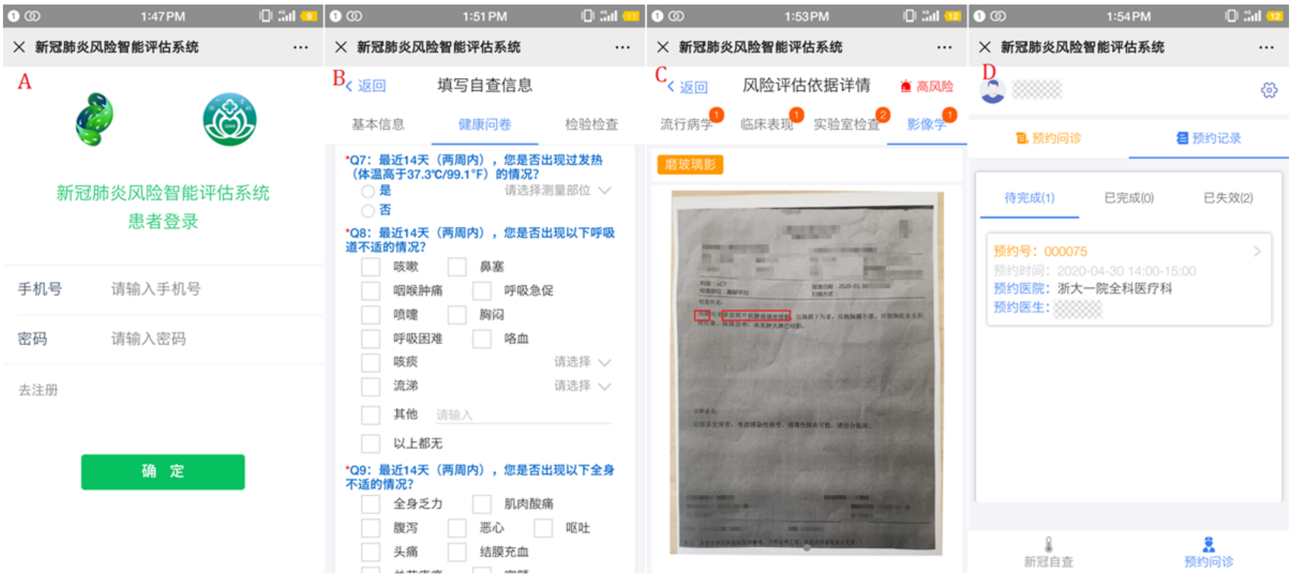
Figure 4. System architecture diagram. API: application programming interface; FHIR: Fast Healthcare Interoperability Resources; HL7: Health Level 7; RESTFUL: representational state transfer.



The mobile app end builds user interface based on the progressive Vue framework. The server end implements data and service interaction with the front end by the representational state transfer application programming interface specification. It is more concise and lighter, both for the processing of uniform resource locators and the encoding of the payload [20]. The server end builds a JavaScript scripting environment by Node.js; improves the system performance using event-driven, nonblocking, and asynchronous input and output models; and optimizes the transmission volume and scale of the app. The data storage adopts the MongoDB open-source database system, which has the characteristics of open-source, cross-platform, and powerful expansion capabilities. To achieve the storage and exchange of patient's clinical data, structured processing data of the laboratory test, and image reports, the system adopts the standard framework of Fast Healthcare Interoperability Resources to ensure the accuracy of clinical data storage and expression. To make full use of the data in the form of pictures contained in the laboratory test and image reports, the system integrates the mature OCR module and NLP module to complete the identification and extraction of relevant clinical information in the picture.

For the patient-end (Figure 5A), the app assists patients in collecting, recording, and transferring their health data. There are two forms of recording health data for patients, including filling in a structured questionnaire by themselves and uploading pictures (Figure 5B and C). The range of health data includes the patient's basic information, a health questionnaire related to COVID-19 early assessment, and clinical test results. In addition, the content of the health questionnaire will be gradually adjusted according to the relevant authoritative guidelines (see details in Multimedia Appendix 1) [15]. In addition, the data of clinical examination, including laboratory test reports such as blood routine test and CRP, and image examination reports such as lung CT reports can be uploaded as pictures directly by the camera of the mobile device. The mobile app communicates with the back end server through a message transmission mechanism. This mechanism establishes a balance between real time data transmission and system availability, which can completely meet the requirements of medical data transmission. After being analyzed by the back end database risk stratification model and related algorithms, the evaluation report will be returned to the patient app, mainly including results and evidence of the risk level evaluation and the corresponding detailed medical advice (Figure 5C). Patients can also update their symptoms and clinical data in real time on the app. The app will give new evaluation results and medical advice immediately. In addition, the patient app supports online patient appointments and provides a list of GPs working in the primary care clinics or fever clinics in designated hospitals (Figure 5D).

Figure 5. The screenshots of DDC19’s patient mobile terminal app.



*Q7：最近14天（两周内），您是否出现过发热（体温高于37.3℃/99.1°F）的情况？

○ 是 请选择测量部位

○ 否

*Q8：最近14天（两周内），您是否出现以下呼吸道不适的情况？

☐ 咳嗽 ☐ 鼻塞

☐ 咽喉肿痛 ☐ 呼吸急促

☐ 喷嚏 ☐ 胸闷

☐ 呼吸困难 ☐ 咯血

☐ 咳痰 请选择

☐ 流涕 请选择

☐ 其他 请输入

☐ 以上都无

*Q9：最近14天（两周内），您是否出现以下全身不适的情况？

☐ 全身乏力 ☐ 肌肉酸痛

☐ 腹泻 ☐ 恶心 ☐ 呕吐

☐ 头痛 ☐ 结膜充血

[redacted]

预约号：000075

预约时间：2020-04-30 14:00-15:00

预约医院：浙大一院全科医疗科

预约医生：[redacted]

For the GP-end (Figure 6A), the app mainly assists GPs in managing their patients, including checking the information of their patients in the system, the results, and the basis of the automatic risk assessment, and giving recommendations based on the assessment results. First, it supports GPs to check their patients according to their risk level or appointment date and searches their patients by their name or mobile phone number (Figure 6B). Second, for the patient who has completed an in-person visit, it supports GPs in uploading the patient’s laboratory test results, image examination reports, and other data to the back end database to which the patient belongs, and they will obtain an update on the patient's risk assessment report. In addition, it also supports GPs in checking the details of their risk assessment results and the previous evaluation records, which helps GPs to grasp the patient's dynamic condition completely and in a timely manner. For a patient who is high risk that did not go to the designated hospital for further investigation timely, the system would send a reminder message for GPs to deal with it as soon as possible. At the same time, GPs can also give recommendations for the patient who applies for offline consultation based on their personal information and submitted clinical data (Figure 6C). GPs can grasp the status of all the residents they managed by these functions of the app. Third, to help GPs in different health institutions grasp the latest prevention and control knowledge of COVID-19 in real time, the GP-end has also developed the function of online training for GPs (Figure 6D).

Figure 6. The screenshots of DDC19’s doctor mobile terminal app.

提交人: [redacted]
更新时间：2020-03-17 13:36:35

提交人: [redacted]
更新时间：2020-03-17 13:32:25

提交人: [redacted]
更新时间：2020-03-17 13:27:35

提交人: [redacted]
更新时间：2020-03-17 13:26:35

提交人: [redacted]
更新时间：2020-03-17 13:19:35

请为患者选择您认为合适的处理意见：

○ 立即定点医院发热门诊就诊
○ 建议全科医生线上或电话咨询
○ 居家观察
○ 其他

请输入

说明文字

Case Studies and Data Validation

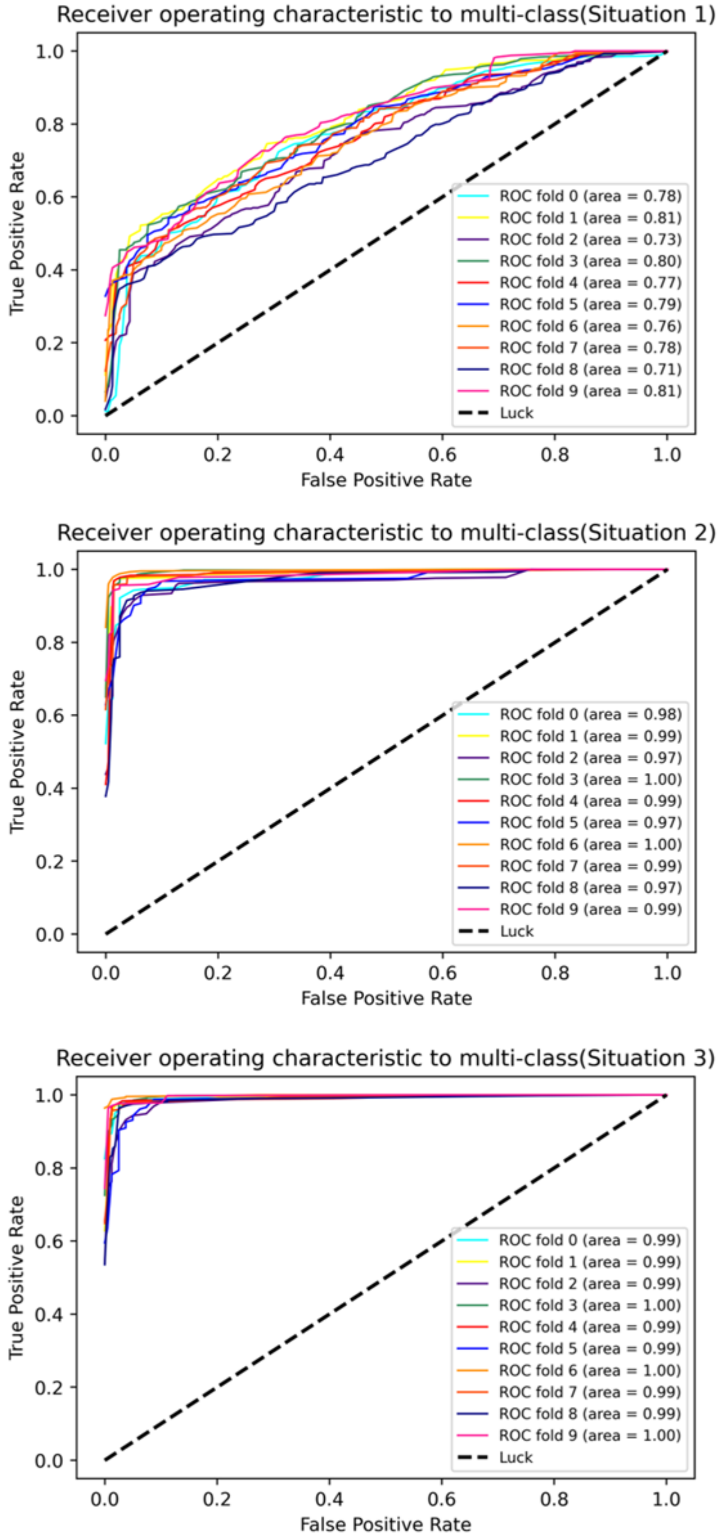
For the purpose of testing and verifying the dynamic risk stratification model of DDC19, 2243 patients’ information were collected from the fever clinic of the first affiliated hospital of Zhejiang University from January 19 to March 11, 2020. The data includes the patients’ basic information (gender and age), chief complaint, medical history, physical examination, laboratory tests, and the lung CT image examination reports. The first affiliated hospital of Zhejiang University is a class A hospital with 2500 beds. In 2019, the number of outpatient and emergency services reached 5 million and 243,300 were discharged.

The patient's medical record number is used as the unique identifier, and the data corresponding to the earliest outpatient visit record of each patient within the range from January 19, 2020, to March 11, 2020, is taken. In terms of data preprocessing, except for age, which is used as a continuous variable, the other data are classified as categorical variables. The imaging examination results were divided into clearly marked “viral pneumonia,” other lung function categories, and no abnormalities. According to the first hospital’s reference threshold, the results of laboratory tests were classified as “lower/normal/higher” or “normal/higher.” Age was standardized, and other missing elements were filled with 0.

In our study, according to the risk assessment model, patients were divided into three groups: high-risk group, moderate-risk group, and low-risk group. Out of the 2243 patients, 628 (28.00%) were in the low-risk group, 1447 (64.51%) were in the moderate-risk group, and 168 (7.49%) were in the high-risk group. Among them, 17 patients were clinically diagnosed with COVID-19; 16 patients were in the high-risk group, and 1 patient was in the moderate-risk group (see details in Multimedia Appendix 2).

To ensure the accuracy of risk stratification and to avoid the model overfitting, we used the three categories of low risk, moderate risk, and high risk as labels, and adopted a 10-fold cross-validation method to evaluate and test the model in different scenarios (different dimensions of personal clinical data accessible at the early stage). When we only used the data of patients’ demographic information, clinical symptoms, and contact history, the data set dimensions were (2243, 15). When the results of blood tests were added, the data set dimensions were (2243, 35). After obtaining the CT imaging results of the patient, the data set dimensions were (2243, 36; see details in Multimedia Appendix 2). Figure 7 demonstrates the training set based on the data set in the previously mentioned three scenarios, labels based on the patient's risk level, the receiver operating characteristic (ROC) curve, and its corresponding area under the curve (AUC) value after 10 cross-validation model test results.

Figure 7. The ROC curve of the dynamic risk stratification model. ROC: receiver operating characteristic.



As can be seen from the ROC curve and its corresponding AUC value, even if only the information of the patient's epidemiology contact history and clinical symptoms were used, the average value of the three classification results of macro-AUC were all above 0.71. When the data of laboratory tests and imaging were added, the macro-AUC increased to above 0.97. Therefore, the model has a good prediction ability for the above three scenarios. The detailed evaluation indicators of the model are shown in Table 2. All indicators were averaged after 10-fold cross-validation, and Class 1, Class 2, and Class 3 refer to low risk, moderate risk, and high risk, respectively.

Table 2. The indicators of the model.

Variables	Situation 1			Situation 2			Situation 3		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Class 1	0.380	0.576	0.456	0.947	0.956	0.951	0.949	0.956	0.952
Class 2	0.750	0.552	0.634	0.976	0.956	0.966	0.980	0.957	0.968
Class 3	0.750	0.947	0.831	0.841	0.941	0.885	0.850	0.982	0.909
Accuracy	0.588	0.588	0.588	0.955	0.955	0.955	0.959	0.959	0.959
Macroaverage	0.627	0.692	0.640	0.921	0.951	0.934	0.926	0.965	0.943
Weighted average	0.646	0.588	0.599	0.958	0.955	0.956	0.961	0.959	0.959

Discussion

Principal Findings

This paper describes a dynamic risk assessment decision support system, which has been used by many Chinese GPs in Zhejiang Province during the COVID-19 outbreak. The DDC19 was designed for GPs working in different situations such as online consultation, assessment evaluation, and triaging in different offline scenarios (community, airport, train station, fever clinic, etc) and following up with suspected patients and discharged patients. It fills in the gap of traditional health care and helps GPs effectively manage residents with different statuses. For patients in DDC19, they can use it to record health data, obtain real time results of risk assessment, and communicate with their GPs without in-person visits; for GPs, they can intuitively grasp their patients' conditions and provide online advice and interventions in real time. DDC19 contributes to the effective triaging of patients, relieves the pressure of offline clinics of designated hospitals to a certain extent, and reduces cross-infection in the hospital during the COVID-19 outbreak.

With the worldwide spread of COVID-19 and the shortage of medical resources, achieving scientific assessment and effectively triaging the patients in different states is the key to control it. Under the actual clinical condition, GPs need to comprehensively evaluate patients offline without specific symptoms based on their epidemiological contact history, symptoms, laboratory, and imaging findings, and advise in conjunction with the guideline's recommendations. Although it can assess patients, distinguish them with different risk levels, and find patients who are at high risk, it has a number of inherent defects that cannot meet the need for efficient prevention and control of COVID-19. At present, there is a lack of mobile medical information systems to meet the needs of patient classification, and it is infeasible to truly complete the triage of patients who are potentially infected with SARS-CoV-2 and other patients with similar symptoms before the outpatient clinic of the designated hospital. Researchers at the University of California, San Diego built a number of COVID-19–related tools to support physician's work based on the EHR system [14]. Although it is beneficial for docking with the process of the original standardized system in the hospital, it cannot directly obtain the data of patients at an early stage and grasp the dynamic conditions of patients. Nowadays, the implementation of early assessment and triage of COVID-19 relies on GP's knowledge and experience. GPs cannot make a dynamic risk assessment for their patients because they cannot completely grasp the condition of patients at each period in a timely manner. In addition, the problem of a large number of panicked patients with mild symptoms or no risk of COVID-19 going to the designated hospitals during the epidemic has not been resolved.

In the traditional model, clinicians need to obtain patient's clinical information within a short time by in-person visits. Due to the limitations of time and environment, it may be arduous to assess the patients' risk for clinicians. In our study, DDC19 can dynamically obtain patients' clinical information by self-report or upload health information, assess their risk level by the dynamic risk stratification logistic regression model, and assist GPs to diagnose and provide recommendations. Therefore, DDC19 helps GPs with online triaging, reduces the pressure of offline clinic, and lowers the risk of cross-infection in the hospital. On the coverage of patient clinical information, our system has been further expanded through questionnaires and inspection reports. The multiclass logistic regression model constructed by the retrospective data (Figure 7 and Table 2) shows that the average recall rate for patients who are high risk reached 0.947, even if the overall accuracy is 0.588 when we only have patients' contact history and clinical symptoms. As we added the results of blood cell counts and CRP, the accuracy rate of the average classification of the model in different risks reached 0.955, and it is close to 0.959 when considering the results of the lung's CT. Therefore, our results proved that the dynamic risk stratification model in DDC19 can accurately identify patients who are high risk by their basic health information (epidemiological history, clinical symptoms, and blood tests). DDC19 can help GPs dynamically manage each patient and expand the scope of COVID-19 risk assessment from traditional outpatient clinics to a new type of continuous health care. In addition, the model proves that only considering the contact history data, clinical symptoms, and blood test data available can achieve the risk stratification effect that combines the results of lung CT. It will reduce the burden of the radiologist and unnecessary waste of health resources.

DDC19 is also a new method for using clinical data, which can respond to emergencies conveniently and quickly. The existing COVID-19–related clinical analysis methods focus on establishing new hypotheses or searching for new research evidence. It is difficult to translate into a direct impact on accurate control and rapid response to the epidemic in a short time. Therefore, we built a complete mobile information system that integrates the workflow in the health care system. On the other hand, although there is a lot of research on informationization and AI-related to COVID-19 at the moment [21-24], most of the reported AI-driven tools are still limited to proof-of-concept models. To the best of our knowledge, there is currently only one article about risk assessment and prehospital triaging of patients with suspected COVID-19 [25], but the tools designed by this study are embedded in EHR systems to provide self-triaging and self-scheduling for patients, so doctor-patient interaction and posthospital follow-ups are not supported. In contrast, the mobile information system we designed fits well with the actual triage practice in business processes and supports GPs' real time grasp and assessment of the patient's health status with better flexibility.

Limitations

This study has some limitations that must be addressed in the next steps of development. First, the complete clinical situation should be considered during the establishment of the dynamic risk assessment model, including the severity of symptoms and history of underlying chronic diseases, but the early retrospective data of fever clinics do not contain these data elements. With further accumulation of relevant clinical data, our system can attempt earlier assessment and triage support methods, and we can also discuss and explore the impact of dynamic changes in patients' clinical information on their COVID-19 risk stratification. Second, as the system is still in the deployment and app stage, the relevant data of patients in the system and in-hospital visits cannot be obtained in a timely manner, so the clinical effects produced by the actual app of the system cannot be evaluated in a timely manner.

Conclusions

DCC19 is a mobile decision support system designed and developed to assist GPs in providing dynamic risk assessments for potential patients during the COVID-19 outbreak. It collects potential patients’ health information by mobile apps and data transmission mechanisms in different situations, assesses their risk levels through a dynamic risk stratification logistic regression model, and helps GPs manage patients and make further clinical decisions.

Acknowledgments

This work was supported by the Major Scientific Project of Zhejiang Lab (2018DG0ZX01), the National Scientific and Technological Major Project of China (2018ZX10715014), “the Fundamental Research Funds for the Central Universities,” and the National Natural Science Foundation of China (81971982, 81771936, 81801796).

Authors' Contributions

This work is based on the cross-cooperation of medical and bioinformatics researchers. JL and JR contributed equally in this work, including conceiving the original idea, designing the whole research process, and reviewing the manuscript. YL and ZW contributed to the design of the research process, data analysis, data interpretation, and wrote the first version of the manuscript. YT, MZ, TZ, and YQ contributed to the administration of the project, data analysis, and data interpretation. KY and YZ collected and cleaned the data. All the authors contributed to the interpretation of the results and the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire of health information.
[DOCX File , 26 KB - jmir_v22i6e19786_app1.docx]

Multimedia Appendix 2
Statistical distribution of patient data characteristics.
[DOCX File , 40 KB - jmir_v22i6e19786_app2.docx]

References

1. Phelan AL, Katz R, Gostin LO. The novel coronavirus originating in Wuhan, China: challenges for global health governance. JAMA 2020 Jan 30:1 [FREE Full text] [doi: 10.1001/jama.2020.1097] [Medline: 31999307]
2. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-03-11]
3. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-04-29]
4. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020 May;109:102433 [FREE Full text] [doi: 10.1016/j.jaut.2020.102433] [Medline: 32113704]

5. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3]
6. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
7. Wu Y, Chen C, Chan Y. The outbreak of COVID-19. J Chin Med Assoc 2020;83(3):217-220. [doi: 10.1097/jcma.0000000000000270]
8. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. Lancet Glob Health 2020 Apr;8(4):e480. [doi: 10.1016/S2214-109X(20)30068-1] [Medline: 32109372]
9. Li DKT, Zhu S. Contributions and challenges of general practitioners in China fighting against the novel coronavirus crisis. Fam Med Com Health 2020 Mar 23;8(2):e000361. [doi: 10.1136/fmch-2020-000361]
10. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
11. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20;26(5):309-313. [doi: 10.1177/1357633x20916567]
12. Jin S, Wang B, Xu H, Lou C, Wait L, Zhao W, et al. AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system in four weeks. medRxiv 2020 Mar 23:1. [doi: 10.1101/2020.03.19.20039354]
13. Du S, Wang J, Zhang H, Cui W, Kang Z, Yang T, et al. Predicting COVID-19 using hybrid AI model. SSRN J 2020 Jan:1. [doi: 10.2139/ssrn.3555202]
14. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
15. The diagnosis and treatment guideline of COVID-19 (Trial Version 7). National Health Commission of the People?s Republic of China. 2020. URL: https://www.cma.org.cn/module/download/downfile.jsp?classid=0&filename=68ff872bf68448e49b553037b3e2fe05.pdf
16. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:1. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
17. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
18. Miao C, Zhuang J, Jin M, Xiong H, Huang P, Zhao Q, et al. A comparative multi-centre study on the clinical and imaging features of comfirmed and uncomfirmed patients with COVID-19. medRxiv 2020 Mar 24:1. [doi: 10.1101/2020.03.22.20040782]
19. Jin X, Lian J, Hu J, Gao J, Zheng L, Zhang Y, et al. Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. Gut 2020 Jun;69(6):1002-1009 [FREE Full text] [doi: 10.1136/gutjnl-2020-320926] [Medline: 32213556]
20. Richards R. Representational state transfer. In: Pro PHP XML and Web Services. Berkeley, CA: Apress; 2006:633-672.
21. Zheng N, Du S, Wang J, Zhang H, Cui W, Kang Z, et al. Predicting COVID-19 in China using hybrid AI model. IEEE Trans Cybern 2020 May:1-14. [doi: 10.1109/tcyb.2020.2990162]
22. Mei X, Lee H, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020 May 19:1. [doi: 10.1038/s41591-020-0931-3] [Medline: 32427924]
23. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 May 11:1. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
24. Ke Y, Peng T, Yeh T, Huang W, Chang S, Wu S, et al. Artificial intelligence approach fighting COVID-19 with repurposing drugs. Biomed J 2020 May 15:1. [doi: 10.1016/j.bj.2020.05.001] [Medline: 32426387]
25. Judson T, Odisho A, Neinstein A, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
COVID-19: coronavirus disease
CRI: credible interval
CRP: C-reactive protein

CT: computed tomography
EHR: electronic health record
GP: general practitioner
NLP: natural language processing
OCR: optical character recognition
ROC: receiver operating characteristic
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.05.20; peer-reviewed by K Araki, C Jacob; comments to author 02.06.20; revised version received 13.06.20; accepted 14.06.20; published 29.06.20.

Please cite as:
Liu Y, Wang Z, Ren J, Tian Y, Zhou M, Zhou T, Ye K, Zhao Y, Qiu Y, Li J
A COVID-19 Risk Assessment Decision Support System for General Practitioners: Design and Development Study
J Med Internet Res 2020;22(6):e19786
URL: http://www.jmir.org/2020/6/e19786/
doi:10.2196/19786
PMID:32540845

Original Paper

Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province

Jitian Li[1*], PhD; Zhe Chen[2*], PhD; Yifei Nie[3], MPH; Yan Ma[1], MPH; Qiaoyun Guo[3], MPH; Xiaofeng Dai[4*], PhD

[1]Henan Provincial Orthopedic Institute, Zhengzhou, China
[2]Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China
[3]Henan Center for Disease Control and Prevention/Henan Key Laboratory of Pathogenic Microorganisms, Zhengzhou, China
[4]Wuxi School of Medicine, Jiangnan University, Wuxi, China
[*]these authors contributed equally

Corresponding Author:
Xiaofeng Dai, PhD
Wuxi School of Medicine
Jiangnan University
Lihu Avenue 1800
Wuxi, 214122
China
Phone: 86 18168870169
Email: 1281423490@qq.com

Abstract

Background: The outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease (COVID-19), has been declared a global pandemic. Identifying individuals whose infection can potentially become severe is critical to control the case fatality rate of COVID-19. However, knowledge of symptoms that are prognostic of COVID-19 severity is lacking.

Objective: The objective of our study was to identify symptoms prognostic of COVID-19 infection severity.

Methods: We analyzed documented symptoms, including fever, cough, fatigue, expectoration, sore throat, chest distress, headache, diarrhea, rhinorrhea, stuffed nose, nausea, vomiting, muscle or joint ache, shortness of breath, and their associations with disease severity using a case series, including 655 confirmed cases from January 23 to February 5, 2020 in Henan Province, China. We also analyzed the influence of individual characteristics, including age, gender, and comorbidities, on symptoms with prognostic value.

Results: Fatigue (95% CI 0.141 to 0.334, *P*<.001), expectoration (95% CI 0.107 to 0.305, *P*<.001) and stuffed nose (95% CI –0.499 to –0.082, *P*=.006) were identified as the prognostic symptoms of COVID-19 patients from the multivariate analysis. Fever occurred in 603/655 (92.1%) of the patients but was not associated with disease severity. Fatigue accounted for 184/655 (28.1%) of the patients and was linearly associated with infection severity with statistical significance. Expectoration occurred in 169/655 (25.8%) patients in the cohort and was the sole prognostic factor for patients with cardiovascular complications, including hypertension. Shortness of breath, chest distress, muscle or joint ache, and dry cough, which occurred in 33 (5%), 83 (12.7%), 78 (11.9%), and 276 (42.1%) of the 655 patients, respectively, were significantly enriched among patients classified as severe. Stuffed nose and nausea were associated with favorable disease severity, especially among male patients. More female than male patients were documented as having muscle or joint ache. Headache was most enriched in patents aged 15 to 39 years, followed by those aged 40 to 64 years, with statistical significance.

Conclusions: Fatigue and expectoration are signs of severe COVID-19 infection. Shortness of breath, chest distress, muscle or joint ache, and dry cough are prevalent in severe patients. Expectoration is commonly present in older individuals and patients with cardiovascular disorders, including hypertension. Shortness of breath is prognostic of severe infection in male patients. Stuffed nose and nausea are favorable prognostic factors of severe infection, especially among male patients.

(J Med Internet Res 2020;22(6):e19636) doi:10.2196/19636

KEYWORDS

prognostic symptoms; COVID-19; severity; CVD; Henan Province

Introduction

In early December 2019, a pneumonia of unknown etiology emerged in Wuhan, a city in China with 11 million permanent residents and 5 million recurrent residents. On December 29, 2019, the first four cases of this pneumonia were reported, all of which were linked to the Huanan Seafood Wholesale Market in Wuhan. On January 7, 2020, a novel coronavirus was identified from the bronchoalveolar lavage fluid of a patient [1] and was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the World Health Organization (WHO) [2]. SARS-CoV-2 is the seventh enveloped RNA coronavirus to be identified [3]; it is transmittable via humans and has a 3-day median incubation time [4]. This virus has rapidly spread worldwide and has become a global health threat [5]. The high binding affinity of SARS-CoV-2 to angiotensin-converting enzyme 2 enables its rapid transmission [6]. Approximately 8 million individuals were infected and over 0.4 million deaths were reported worldwide as of June 2020 [7]. The death rate varies among countries and reached as high as 27.1% in Yemen [7].

Suspected cases were identified as having “fever or respiratory symptoms” and “traveling history or contact with confirmed infections within 2 weeks” [8]. Unlike the symptoms of severe acute respiratory syndrome (SARS), where fever was the symptom in approximately 100% of infected individuals [9], fever is only observed in 87.9% of patients with COVID-19 on admission [4]; other symptoms, such as cough (67.7%) and fatigue (38.1%), also frequently occur in SARS-CoV-2–infected individuals. Despite the many reports on symptoms associated with COVID-19, little effort has been devoted to the identification of symptoms associated with its severity. This study aims to identify symptoms with prognostic value on disease severity and their correlations with individual characteristics such as age, gender, and comorbidities to aid the prognosis of COVID-19 severity.

Methods

Data Source

This case series was collected by the Center for Disease Control and Prevention of Henan Province (Henan CDC) from 279 hospitals in the province. It includes 655 confirmed patients with COVID-19 who showed symptoms on admission and were admitted to hospitals from January 23, 2020, to February 5, 2020, with February 5 being the last follow-up date. Oral consent was obtained from the patients. All enrolled patients were diagnosed according to the WHO interim guidance [10].

Epidemic, clinical, and severity data were obtained with data collection forms from electronic medical records as part of standard care. The information recorded included demographic data, comorbidities, symptoms, and chest computed tomography (CT) scans. The date of disease onset was defined as the day when the symptom was noted. This study was approved by the ethics commissions of the Henan CDC with a waiver of informed consent. Cardiovascular disease (CVD) is a group of disorders of the heart and blood vessels, including cerebrovascular disease, coronary heart disease, cerebrovascular disease, peripheral arterial disease, and rheumatic heart disease.

We stratified severity into three groups: light, normal, and severe. According to the Novel Coronavirus Diagnostic and Therapeutic Plan (Seventh Edition) [11], COVID-19 severity was initially divided into four types: light, normal, severe, and terminal. We merged severe and terminal patients into one group, named the severe group. The clinical symptoms of patients in the light group were mild, with no pneumonia found in imaging. Patients in the normal group had fever, respiratory tract infection, and other symptoms, with manifestation of pneumonia observable in imaging. In adults, if one of the following three conditions was satisfied, the patient was classified as severe: 1) shortness of breath, respiratory rate ≥30 times per minute; oxygen saturation ≤93% in the resting state; arterial blood sample partial pressure (PaO_2)/oxygen concentration (FiO_2) ≤300 millimeters of mercury. Among children, if any of the following criteria were met, the child was classified as severe: shortness of breath (≤2 months of age, respiratory rate ≥60 times/min; 2 to 12 months of age, respiratory rate ≥50 times/min; 1-5 years of age, respiratory rate ≥40 times/min; >5 years of age, respiratory rate ≥30 times/min), excluding the effects of fever and crying; oxygen saturation in the resting state is ≤92%; assisted respiration (moaning, alar fluttering, three-concave sign); cyanosis; intermittent apnea; drowsiness; convulsion; refusing to eat or difficulty in feeding; and signs of dehydration. Patients with critically severe cases satisfied the following criteria: respiratory failure that requires mechanical ventilation; shock; other organ failure that requires intensive care unit monitoring and treatment.

Laboratory Testing

Throat or nose swab samples were collected from patients suspected of having SARS-CoV-2 infection for total RNA extraction using the respiratory sample RNA isolation kit (Shanghai BioGerm Medical Technology Co Ltd, Catalog No. ZC-HX-201-2), followed by real time reverse transcription–polymerase chain reaction (RT-PCR) using a SARS-CoV-2 nucleic acid detection kit (Shanghai BioGerm Medical Technology Co Ltd) in the biosafety level 2 lab at Henan CDC. Targeting the open reading frame (ORF1a/b), the primers and sequences were forward primer CCCTGTGGGTTTTACACTTAA, reverse primer ACGATTGTGCATCAGCTGA, and probe 5'-FAM-CCGTCTGCGGTATGTGGAAAGGTTATGG-BHQ1-3'. Targeting nucleocapsid protein, the primers and sequences were forward primer GGGGAACTTCTCCTGCTAGAAT, reverse primer CAGACATTTTGCTCTCAAGCTG, and probe 5'-FAM-TTGCTGCTGCTTGACAGATT-TAMRA-3'. Conditions for the amplifications were 50 °C for 10 minutes and 95 °C for 5 minutes, followed by 40 cycles of 95 °C for 10 seconds and 55 °C for 40 seconds. Following recommendations by the Chinese National Institute for Viral Disease Control and Prevention [12], positive and negative tests were defined as cycle threshold (C_t) values <37 and ≥40, respectively; samples with C_t values between these thresholds were subjected to retesting. A case was confirmed if two targets (ORF1a or 1b, nucleocapsid protein) tested positive by real-time RT-PCR in the initial test or both tests (when a retest was needed).

Statistical Analysis

Continuous variables were described using mean, SD, median, interquartile range (IQR), and range, and categorical variables were described by frequency and percentage. Means for continuous variables were compared using independent group tests when the data were normally distributed (Shapiro-Wilk test); otherwise, the Kruskal-Wallis H test was used (adjusted by Bonferroni correction). Proportions for categorical variables were compared using the chi-square test or Fisher exact test. The correlation of two variables was compared using Spearman rank correlation. The influencing variables for the severity of COVID-19 were analyzed using a linear regression model (forward method). All statistical analyses were performed using SPSS version 23.0 (IBM Corp). A 2-sided α less than .05 was considered statistically significant.

IT Infrastructure

We implemented the new Public Health Emergency Management Information System of Henan Province and extracted data from the system to analyze the epidemiological characteristics of COVID-19 patients in Henan Province and develop a COVID-19 cluster statistical information template. In addition, the hospital information system, dashboards, electronic prescription system, and cloud-based medical image sharing system were used to facilitate the analysis of the clinical data of COVID-19 patients.

Results

In this case series including 655 COVID-19 patients, regarding disease severity, 163 (24.9%) patients were classified as light, 420 (64.1%) as normal, and 72 (11%) as severe (Table 1). Of the infected individuals, 12/655 (1.8%) were less than 15 years of age, 265 (40.5%) were between 15 and 39 years of age, 322 (49.2%) were between 40 and 64 years of age, and 71 (8.5%) were aged 65 years or older (Table 1). Most patients were male (367/655, 56.0%). Of the 150/655 patients (22.9%) who had coexisting medical disorders, 89 (59.3%) had CVD. Given the high percentage of CVD comorbidities in this dataset, we analyzed symptoms prognostic of COVID-19 severity separately among patients with CVD, patients without CVD, and patients without comorbidities. The average number of days from illness onset to diagnosis was 5.66 (SD 3.64, range 0-32). Among the 655 patients included in this case series, 634 (96.8%) had chest CT scans, and 535 (84.4%) showed typical pneumonia features (bilateral ground glass opacities, Figure 1).

Among all documented symptoms in this case series, fever and dry cough were prevalent in all patient cohorts as stratified by COVID-19 severity. Fatigue and expectoration were enriched symptoms in the severe and normal groups (Figure 2A).

The symptom of fever occurred in 603/655 (92.1%) of the patients (Table 1) but was not associated with disease severity. A higher “highest temperature of fever” was typically associated with more severe SARS-CoV-2 infection; however, it is difficult to precisely define high and low highest temperatures of fever given the high dependence of this parameter on the time slot when it was measured and high individual heterogeneity among patients. The mean highest temperature in the light group was 38.0ºC (range 37.6-38.5ºC), that in the normal group was 38.0ºC (range 37.8-38.5ºC), and that in the severe group was 38.4ºC (range 38.0-38.9ºC) (χ^2_2=15.5, P<.001).

Other symptoms, including dry cough, fatigue, expectoration, chest distress, muscle or joint ache, shortness of breath, and multiple symptoms, all convey significant prognostic value on disease severity. Fatigue was the most prevalent symptom (184/655, 28.1%) among patients (Table 1) and could significantly (χ^2_2=20.8, P<.001) stratify and linearize patients into light, normal, and severe groups regarding clinical severity (Table 2, Figure 2A). Expectoration was observed in 169/655 (25.8%) of the patients, with significant (χ^2_2=14.5, P=.001) prognostic value on disease severity (Table 1, Table 2). Expectoration significantly (χ^2_2=6.6, P=.04) differed among patients with CVD and patients with complications other than CVD as well as patients without coexisting disorders (Multimedia Appendix 1), and it was the sole explanatory symptom included in the linear regression among CVD patients (β=0.310, 95% CI 0.042-0.579, P=.02, Table 3). Shortness of breath was significantly (χ^2_2=18.3, P<.001) enriched in the severe group (Table 2) and was observed in 33/655 (5.0%) of the patients infected with COVID-19 (Table 1). Chest distress occurred in 12.7% of this cohort (Table 1), and the sample group was significantly enriched with severely infected patients (χ^2_2=11.367, P=.003, Table 2). Muscle or joint ache was documented in 78/655 (11.9%) of the patients and was significantly enriched in severe patients (χ^2_2=7.7, P=.02, Table 2) and female patients (χ^2_1=4.5, P=.03, Multimedia Appendix 1). Dry cough, although it could stratify patients according to COVID-19 severity with statistical significance (χ^2_2=7.4, P=.03, Table 2), did not show good discrimination power (Figure 2B); this may be due to the high prevalence of dry cough among patients (276/655, 42.1%, Table 1). Headache was enriched in the 15 to 39 years age group, followed by the 40 to 64 years age group, with statistical significance (χ^2_3=11.6, P=.009, Multimedia Appendix 1); however, headache could not differentiate disease severity. The majority of the patients (501/655, 76.5%) reported multiple symptoms (Table 1).

Table 1. Documented symptoms and general clinical characteristics of 655 patients with COVID-19 (N=655).

Characteristic	n (%)
Disease severity	
Light	163 (24.9)
Normal	420 (64.1)
Severe	72 (11)
Symptoms	
Fever (°C)	
<37.3	23 (3.5)
37.3-38.0	287 (43.8)
38.1-39.0	241 (36.8)
>39.0	37 (5.6)
Temperature not documented	15 (2.1)
Dry cough	276 (42.1)
Fatigue	184 (28.1)
Expectoration	169 (25.8)
Chest distress	83 (12.7)
Headache	80 (12.2)
Muscle or joint ache	78 (11.9)
Sore throat	70 (10.7)
Rhinorrhea	41 (6.3)
Shortness of breath	33 (5.0)
Diarrhea	33 (5.0)
Stuffed nose	30 (4.6)
Nausea	22 (3.4)
Vomiting	19 (2.9)
Other symptoms	7 (1.1)
Multiple symptoms	501 (76.5)
Age (years)	
<15	12 (1.8)
15-39	265 (40.5)
40-64	322 (49.2)
≥65	56 (8.5)
Gender	
Male	367 (56.0)
Female	288 (44.0)
Coexisting disorders	
CVD[a]	89 (59.3)
Disorders other than CVD	61 (40.7)
None	505 (77.1)

[a] CVD: cardiovascular disease (including hypertension).

Figure 1. Chest computed tomographic images of a female patient aged 56 years infected with coronavirus disease. The images show ground glass opacity in both lungs on day 5 after symptom onset.

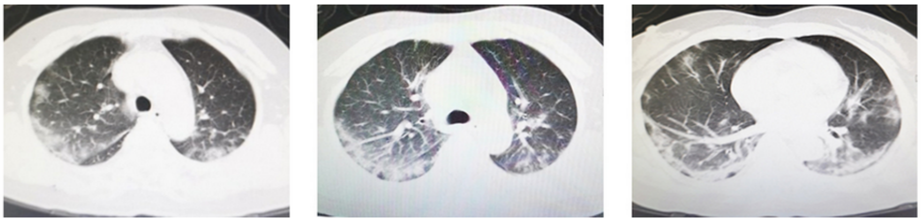


Figure 2. Summarized report of documented symptoms. Plots of (A) patient number of documented symptoms and (B) patient percentage of prognostic symptoms in cohorts stratified by coronavirus disease severity.

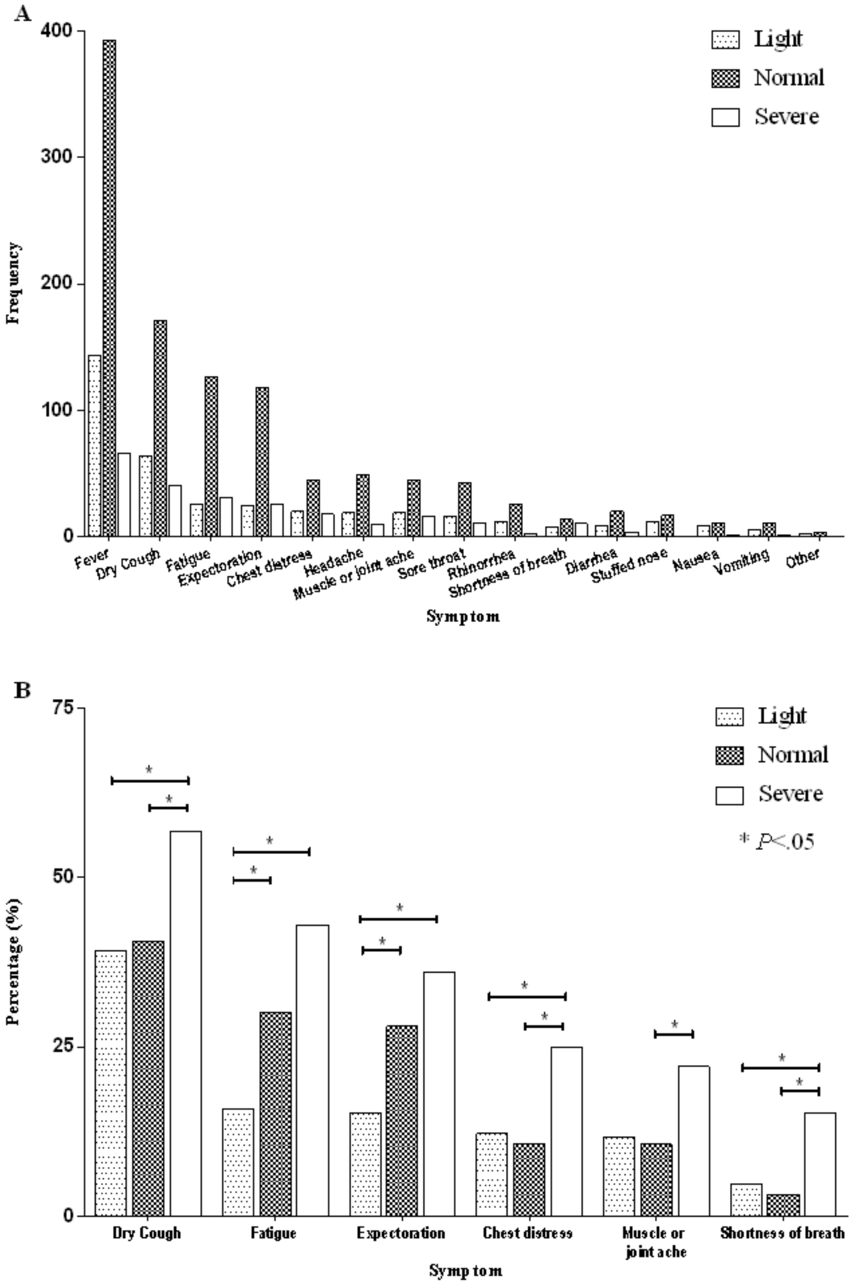


Table 2. Symptoms stratified by disease severity among patients with COVID-19 (N=655).

Symptom	Light (n=163), n (%)	Normal (n=420), n (%)	Severe (n=72), n (%)	Chi-square (*df*)	*P* value
Fever					
All fever	144 (88.3)	393 (93.6)	66 (91.7)	4.4 (*2*)	.11
<37.3ºC	4 (2.9)	18 (4.6)	1 (1.6)	21.5 (*6*)[b]	.002
37.3-38.0ºC	80 (58.4)[a]	187 (48.1)	20 (32.3)[a]		
38.1-39.0ºC	43 (31.4)[a]	166 (42.7)	32 (51.6)[a]		
>39.0ºC	10 (7.3)	18 (4.6)[a]	9 (14.5)[a]		
Temperature not documented	7 (4.3)	4 (0.1)	4 (5.6)		
Dry cough	64 (39.3)	171 (40.7)	41 (56.9)[c]	7.4 (*2*)	.03
Fatigue	26 (16.0)[c]	127 (30.2)	31 (43.1)	20.8 (*2*)	<.001
Expectoration	25 (15.3)[c]	118 (28.1)	26 (36.1)	14.5 (*2*)	.001
Chest distress	20 (12.3)	45 (10.7)	18 (25.0)[c]	11.4 (*2*)	.003
Headache	20 (12.3)	49 (11.7)	10 (13.9)	0.3 (*2*)	.86
Muscle or joint ache	19 (11.7)	45 (10.7)[a]	16 (22.2)[a]	7.7 (*2*)	.02
Sore throat	16 (9.8)	43 (10.2)	11 (15.3)	1.8 (*2*)	.41
Rhinorrhea	12 (7.4)	26 (6.2)	3 (4.2)	0.9 (*2*)	.65
Shortness of breath	8 (4.9)	14 (3.3)	11 (15.3)[c]	18.3 (*2*)	<.001
Diarrhea	9 (5.5)	20 (4.8)	4 (5.6)	0.2 (*2*)	.91
Stuffed nose	12 (7.4)	17 (4.0)	1 (1.4)	4.8 (*2*)	.09
Nausea	9 (5.5)	11 (2.6)	2 (2.8)	3.1 (*2*)	.21
Vomiting	6 (3.7)	11 (2.6)	2 (2.8)	0.7 (*1*)	.78
Other	3 (1.8)	4 (1.0)	0 (0)	1.3 (*1*)	.49
Multiple symptoms	111 (68.1)	327 (77.9)	63 (87.50)	11.7 (*2*)	.003

[a] Pairwise significance.

[b] Calculated with the R by C chi-square test.

[c] Significant compared with the other two groups.

Linear models were established to explain disease severity by including all patients or cohort groups stratified by age, gender, and coexisting disorder. All models constructed were significant (Table 3). Fatigue was the most frequently present symptom in these equations, followed by expectoration, and the coefficients of both symptoms were positive (Table 3). While expectoration was the sole symptom associated with patients with CVD, fatigue was linked to patients with coexisting disorders other than CVD (Table 3). Excluding patients with comorbidities did not change the variables included in the model except for slight variations in the coefficients. Stuffed nose was a negative explanatory variable in the models including all patients, individuals without comorbidities, and male patients only (Table 3). Nausea was a negative explanatory variable and played a dominant role in the equation established for male patients (Table 3).

The symptoms of highest temperature of fever (correlation 0.166, 95% CI 0.082-0.251, *P*<.001), expectoration (correlation 0.104, 95% CI 0.022-0.184, *P*=.012) and shortness of breath (correlation 0.125, 95% CI 0.001-0.041, *P*=.002) were significantly correlated with days from illness onset to diagnosis (Table 4).

Table 3. Linear models of clinical severity as explained by symptoms.

Stratification factor	Parameter from stepwise modeling	Regression coefficient	95% CI	*P* value	Model[a]
All	Constant	1.755	1.697 to 1.812	<.001	Y=0.237*F+0.206*E-0.291*S+1.755
	Fatigue	0.237	0.141 to 0.334	<.001	
	Expectoration	0.206	0.107 to 0.305	<.001	
	Stuffed nose	0.291	–0.499 to –0.082	.006	
Age	Constant	1.499	1.272 to 1.725	<.001	
≥65 years	Expectoration	0.656	0.253 to 1.059	.002	Y=0.460*F+0.656*E+1.499
	Fatigue	0.460	0.066 to 0.854	.02	
	Constant	1.820	1.739 to 1.900	<.001	
40-64 years	Fatigue	0.202	0.072 to 0.331	.002	Y=0.202*F+0.137*E+1.82
	Expectoration	0.137	0.001 to 0.274	.048	
	Constant	1.789	1.710 to 1.868	<.001	
15-39 years	Fatigue	0.166	0.008 to 0.324	.04	Y=0.166*F+1.789
	Constant	1.20	0.918 to 1.482	<.001	
<15 years	Fatigue	0.800	0.110 to 1.490	.03	Y=0.8*F+1.2
Gender	Constant	1.783	1.706 to 1.860	<.001	
Male gender	Fatigue	0.219	0.091 to 0.346	.001	Y=0.219*F+0.212*E-0.513*N-0.343*S+0.276*B+1.783
	Expectoration	0.212	0.083 to 0.340	.001	
	Nausea	–0.513	–0.880 to –0.147	.006	
	Stuffed nose	–0.343	–0.602 to –0.084	.01	
	Shortness of breath	0.276	0.027 to 0.525	.03	
	Constant	1.718	1.632 to 1.803	<.001	
Female gender	Fatigue	0.252	0.100 to 0.404	.001	Y=0.252*F+0.164*E+1.718
	Expectoration	0.164	0.006 to 0.322	.04	
Coexisting disorder	Constant	1.864	1.727 to 2.000	<.001	
Coexisting CVD[b]	Expectoration	0.310	0.042 to 0.579	.02	Y=0.31*E+1.864
	Constant	1.725	1.548 to 1.902	<.001	
Coexisting disorder other than CVD	Fatigue	0.418	0.116 to 0.720	.008	Y=0.418*F+1.725
	Constant	1.736	1.672 to1.800	<.001	
No coexisting disorder	Fatigue	0.264	0.152 to 0.376	<.001	Y=0.264*F+0.227*E-0.314*S+1.736
	Expectoration	0.227	0.111 to 0.343	<.001	
	Stuffed nose	–0.314	–0.548 to –0.080	.009	

[a] B: shortness of breath. E: expectoration. F: fatigue. N: nausea. S: stuffed nose.

[b] CVD: cardiovascular disease.

Table 4. Spearman correlation analysis of symptoms and days from illness onset to diagnosis.

Symptoms	Days from illness onset to diagnosis		
	Correlation	95% CI	*P* value
Fever	–0.039	–0.119 to 0.065	.35
Highest temperature of fever	0.166[a]	0.082 to 0.251	<.001
Temperature <37.3ºC	–0.130	–0.492 to 0.306	.55
Temperature 37.3-38.0ºC	0.006	–0.116 to 0.124	.91
Temperature 38.1-39.0ºC	0.009	–0.121 to 0.139	.89
Temperature >39.0ºC	0.109	–0.190 to 0.403	.52
Dry cough	0.058	–0.021 to 0.132	.15
Fatigue	0.023	–0.067 to 0.101	.58
Expectoration	0.104[b]	0.022 to 0.184	.012
Chest distress	0.070	–0.011 to 0.157	.09
Headache	–0.037	–0.117 to 0.044	.37
Muscle or joint ache	0.013	–0.069 to 0.093	.76
Sore throat	0.008	–0.071 t 0.080	.85
Rhinorrhea	–0.041	–0.114 to 0.033	.33
Shortness of breath	0.125[a]	0.001 to 0.041	.002
Diarrhea	–0.004	–0.086 to 0.087	.92
Stuffed nose	–0.065	–0.152 to 0.033	.12
Nausea	0.018	–0.072 to 0.103	.66
Vomiting	–0.032	–0.131 to 0.063	.43
Multiple symptoms	0.079	–0.002 to 0.158	.056

[a] *P*<.01 (two-tailed).

[b] *P*<.05 (two-tailed).

Discussion

Principal Findings

The main findings of this study were that fatigue and expectoration are signs of severe COVID-19 infection and that stuffed nose and nausea are favorable prognostic factors of disease severity.

Two symptoms, fatigue and expectoration, showed linear associations with COVID-19 severity (Figure 2B). Four symptoms, namely shortness of breath, chest distress, muscle or joint ache, and dry cough, were more commonly present in severe patients. Thus, if more patients are documented with these symptoms, more resources for intensive medical care should be administrated.

These prognostic symptoms were interconnected with gender, CVD, and age. The enrichment of shortness of breath in severe patients was the most evident, especially among male patients; this suggests its feasibility as a sign (with relatively low type I error) for intensive care among confirmed cases. Stuffed nose and nausea were associated with less severe COVID-19, especially among male patients. Expectoration was significantly associated with CVD complications, suggesting a correlation between CVD and lower respiratory tract infection. An age of 40 years was shown to be a breakpoint for symptoms prognostic of disease severity. While fatigue conveyed prognostic value for all age groups, expectoration showed significance when the patients' age exceeded 40 years and dominated the model when their age exceeded 65 years. This may be due to the increased likelihood of developing CVD with increasing age and explainable by the strong association between expectoration and CVD comorbidity. As a clinical suggestion, individuals concomitantly having these characteristics and symptoms should be suspected for infection and given immediate quarantine and potential intensive medical care during the COVID-19 epidemic.

It is known that immune response to virus infection plays a vital role in the inflammation involved in heart diseases such as myocarditis, atherosclerosis, and cardiac insufficiency, and it constitutes the pathogenesis of cardiac disorder in humans [13]. Shortness of breath is a typical sign of heart and lung conditions, and fever is a symptom of stimulated immune response to infection. It was observed from our study that expectoration was associated with CVD and male gender; expectoration, shortness of breath, and fever were significantly correlated with days from illness onset to diagnosis. Therefore, we expect that patients with comorbidities, especially CVD, would experience exacerbated COVID-19 severity.

It is worth mentioning that symptoms prognostic of COVID-19 severity differ from symptoms for early COVID-19 diagnosis. While fever and dry cough were the most prevalent symptoms among infected individuals, fatigue, expectoration, and chest distress conveyed prognostic value on disease severity. This can be explained by the small percentage of patients represented in the whole case series (72, 11%) who could not be predicted from prevalent symptoms.

Although many case series including higher numbers of cases than that in this study have been reported [4,14-16], given the rapid changing global situation of COVID-19 and the sharp rise in the number of infected cases during the past few months, relatively little research has been devoted to studying the prognostic value of symptoms on severity. The sole relevant study was reported by Dong et al [17], who analyzed data from 663 patients. Compared with Dong's study, which analyzed two symptoms (expectoration and muscle ache), one laboratory test index (albumin), and one patient characteristic (gender), our study concentrates on a more focused and complete list or characteristics that encompasses 14 symptoms. However, this study is limited by its high dependence on the accuracy and completeness of the symptoms recorded for each patient.

Conclusions

Our study provides a statistical analysis of documented symptoms of 655 confirmed COVID-19 patients from Henan Province to aid SARS-CoV-2 diagnosis and prognosis. We conclude that fatigue and expectoration are the most important symptoms prognostic of severe COVID-19, and gender, age, and CVD comorbidity are factors associated with these symptoms; muscle or joint ache commonly occurs in female patients, and younger patients are likely to develop headache; and high temperature in fever, expectoration, and shortness of breath are typically associated with delayed diagnosis.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant No. 81972789), the National Science and Technology Major Project (Grant No. 2018ZX10302205-004-002), and the Major Project of TCM Research in Henan Province (Grant No. 2018ZYZD01).

Authors' Contributions

DXF is the principal investigator and guarantor of the paper, had full access to all the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis. DXF and LJT obtained the funding. DXF designed the research and drafted the manuscript. LJT and MY performed the statistical analysis. DXF, LJT, CZ, NYF, MY, and GQY contributed to the acquisition, analysis, or interpretation of data and critically reviewed and revised the article for important intellectual content. All authors approved the final manuscript and decided to submit the article for publication. LJT, ZC, and XFD contributed equally as co-first authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Symptoms stratified by CVD disorder, gender, and age in a case series of COVID-19 patients from Henan Province (N=655).
[DOCX File , 25 KB - jmir_v22i6e19636_app1.docx]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017] [Medline: 31978945]
2. Guo Y, Cao Q, Hong Z, Tan Y, Chen S, Jin H, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020 Mar 13;7(1):11 [FREE Full text] [doi: 10.1186/s40779-020-00240-0] [Medline: 32169119]
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5] [Medline: 31986264]
4. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
5. Novel Coronavirus (2019-nCoV) Situation Report-15. World Health Organization. 2020 Feb 04. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200204-sitrep-15-ncov.pdf?sfvrsn=88fe8ad6_4 [accessed 2020-02-15]
6. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020 Mar 13;367(6483):1260-1263 [FREE Full text] [doi: 10.1126/science.abb2507] [Medline: 32075877]

7. Coronavirus Disease (COVID-19) Situation Report – 150. World Health Organization. 2020 Jun 18. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200618-covid-19-sitrep-150.pdf?sfvrsn=aa9fe9cf_2 [accessed 2020-06-19]
8. New Coronavirus Infection Pneumonia Prevention and Control Plan (Second Edition). Webpage in Chinese. National Health Commission of the People's Republic of China. 2020 Jan 22. URL: http://www.nhc.gov.cn/jkj/s3577/202001/c67cfe29ecf1470e8c7fc47d3b751e88.shtml [accessed 2020-02-15]
9. Peiris J. Severe Acute Respiratory Syndrome (SARS). J Clin Virol 2003 Dec;28(3):245-247 [FREE Full text] [doi: 10.1016/j.jcv.2003.08.005] [Medline: 14522062]
10. Patel A, Jernigan DB, 2019-nCoV CDC Response Team. Initial Public Health Response and Interim Clinical Guidance for the 2019 Novel Coronavirus Outbreak - United States, December 31, 2019-February 4, 2020. MMWR Morb Mortal Wkly Rep 2020 Feb 07;69(5):140-146 [FREE Full text] [doi: 10.15585/mmwr.mm6905e1] [Medline: 32027631]
11. Novel coronavirus diagnostic and therapeutic plan (Seventh Edition). Webpage in Chinese. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989/files/ce3e6945832a438eaae415350a8ce964.pdf [accessed 2020-06-11]
12. Specific primers and probes for detection 2019 novel coronavirus. Webpage in Chinese. Chinese National Institute for Viral Disease Control and Prevention. 2020 Jan 21. URL: http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html [accessed 2020-06-18]
13. Sengur A. An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases. Comput Biol Med 2008 Mar;38(3):329-338. [doi: 10.1016/j.compbiomed.2007.11.004] [Medline: 18177849]
14. Korean Society of Infectious Diseases, Korean Society of Pediatric Infectious Diseases, Korean Society of Epidemiology, Korean Society for Antimicrobial Therapy, Korean Society for Healthcare-associated Infection Control and Prevention, Korea Centers for Disease Control and Prevention. Report on the Epidemiological Features of Coronavirus Disease 2019 (COVID-19) Outbreak in the Republic of Korea from January 19 to March 2, 2020. J Korean Med Sci 2020 Mar 16;35(10):e112 [FREE Full text] [doi: 10.3346/jkms.2020.35.e112] [Medline: 32174069]
15. Hao S, Zhang S, Lian J, Jin X, Ye C, Cai H, et al. Liver Enzyme Elevation in Coronavirus Disease 2019: A Multicenter, Retrospective, Cross-Sectional Study. Am J Gastroenterol 2020 Jun 01:preprint. [doi: 10.14309/ajg.0000000000000717] [Medline: 32483005]
16. Cariou B, Hadjadj S, Wargny M, Pichelin M, Al-Salameh A, Allix I, CORONADO investigators. Phenotypic characteristics and prognosis of inpatients with COVID-19 and diabetes: the CORONADO study. Diabetologia 2020 May 29:preprint [FREE Full text] [doi: 10.1007/s00125-020-05180-x] [Medline: 32472191]
17. Zhang J, Wang X, Jia X, Li J, Hu K, Chen G, et al. Risk factors for disease severity, unimprovement, and mortality in COVID-19 patients in Wuhan, China. Clin Microbiol Infect 2020 Jun;26(6):767-772 [FREE Full text] [doi: 10.1016/j.cmi.2020.04.012] [Medline: 32304745]

Abbreviations

COVID-19: coronavirus disease
C_t: cycle threshold
CT: computed tomography
CVD: cardiovascular disease
Henan CD: Center for Disease Control and Prevention of Henan Province
IQR: interquartile range
RT-PCR: real time reverse transcription–polymerase chain reaction
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.04.20; peer-reviewed by P Banik, E van Gorp, A Bhagavathula; comments to author 06.06.20; revised version received 14.06.20; accepted 15.06.20; published 30.06.20.

Please cite as:
Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X
Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province
J Med Internet Res 2020;22(6):e19636
URL: https://www.jmir.org/2020/6/e19636
doi:10.2196/19636
PMID:32544071

Original Paper

Perceptions About Technologies That Help Community-Dwelling Older Adults Remain at Home: Qualitative Study

Henk Verloo[1,2*], PhD; Thomas Kampel[3*], MscN; Nicole Vidal[4*], MD; Filipa Pereira[1*], PhD

[1]School of Health Sciences, HES-SO Valais/Wallis, Sion, Switzerland
[2]University Hospital Lausanne, Service of Old Age Psychiatry, Prilly, Switzerland
[3]La Source, School of Nursing Sciences, HES-SO, University of Applied Sciences and Arts Western Switzerland, Lausanne, Switzerland
[4]Conseil Départemental de la Haute Savoie, Annecy, France
[*]all authors contributed equally

Corresponding Author:
Henk Verloo, PhD
School of Health Sciences
HES-SO Valais/Wallis
5, Chemin de l'Agasse
Sion, 1950
Switzerland
Phone: 41 276036752
Email: henk.verloo@netplus.ch

Abstract

Background: The population of Europe is aging rapidly. Most community-dwelling older adults (CDOAs) want to remain in their homes, particularly those experiencing functional decline. Politicians and academics repeatedly praise technological instruments for being the preferred solution for helping older adults with deteriorating health to remain at home.

Objective: This study aimed to understand the perceptions of CDOAs and their informal caregivers (ICs) and professional caregivers (PCs) about technologies that can help keep older adults at home.

Methods: This qualitative study used personal interviews, focus groups, and photo-elicitation interviews to better understand the perceptions of a convenience sample of 68 CDOAs, 21 ICs, and 32 PCs.

Results: A fraction of CDOAs did not perceive technological instruments to be a very useful means of helping them remain at home. However, the ICs and PCs were more positive. The CDOAs preferred and were more willing to adopt technologies related to their mobility and safety and those that would help slow down their cognitive decline. The ICs preferred technological aids that assist in the activities of daily living as well as safety-related technologies for detecting falls and helping to locate disoriented older adults. The PCs preferred integrated communication and information systems to improve collaboration between all stakeholders, housing equipped with technologies to manage complex care, high-performance ancillary equipment to transfer people with reduced mobility, and surveillance systems to ensure safety at home.

Conclusions: Although our study reports that CDOAs have limited interest in innovative technologies to help them remain at home, their technological skills will undoubtedly improve in the future, as will those of ICs and PCs. Technological tools will play an increasingly important role in home health care.

(J Med Internet Res 2020;22(6):e17930) doi:10.2196/17930

KEYWORDS

technology; gerontechnology; photo-elicitation; informal caregivers; cognitive impairment; professional caregivers; interviews; focus groups; content analysis; physical impairment; frailty

Introduction

Background

The population of Europe is aging rapidly [1]. Most countries will have to cope with increasing numbers of frail, community-dwelling older adults (CDOAs) who are losing their autonomy and becoming dependent on assistance [1]. Despite their disabilities, 9 out of 10 CDOAs want to remain at home, even those experiencing significant functional decline and loss of autonomy [2]. In this context, caring for CDOAs has become a major health care, social, economic, and political issue [3]. Demographic transition results in an aging society, with fewer young adults available to support the needs of a dramatically rising number of dependent older adults [4]. The costs of specialist home care and social care to support CDOAs will increase exponentially. Innovation in the management of person-centered care is unavoidable, and it must consider medical history, life expectancy, economic constraints, and society's expectations for optimal autonomy and mobility [5,6]. Technology can provide proactive solutions for some of the main health and societal problems encountered during aging, such as loss of independence, chronic diseases, psychopathological disorders, and falls at home [6,7]. Despite the omnipresence of technology in modern society, research into its use for improving the daily lives of frail, cognitively impaired CDOAs is limited in naturalistic settings. Technologies can offer innovative ways of improving the health and quality of life of frail CDOAs—or other people losing their autonomy—and help them to remain at home [6].

Technology aimed at older adults is usually termed gerontechnology (which this paper will use as a synonym of technology). It is destined to have an important future role in providing support and solutions, monitoring health status to optimize autonomy, and improving the quality of life of CDOAs, regardless of their level of dependency [8,9]. Gerontechnologies usually have two main, and in most cases, complementary purposes: (1) strengthening the (objective) monitoring of older adults by remotely collecting large amounts of data to alert their informal caregivers (ICs) or professional caregivers (PCs) about health decompensation, so that they can anticipate and implement appropriate intervention strategies, and (2) remotely intervening with CDOAs for the first interaction (using voice or images) [10].

Cornet and Carre [11] mentioned that certain technologies could help maintain levels of autonomy despite debilitating age-related diseases, improve social cohesion, reduce loneliness, compensate for declining capacities, help both ICs and PCs, and, finally, reduce some of the effects of pathological aging. Science and technology will also serve CDOAs by ensuring safer environments, promoting and strengthening their independence and quality of life, and supporting both their ICs and PCs [11-13]. New devices should make it possible to alert ICs and PCs about the risks of health decompensation so that they can anticipate and implement appropriate intervention strategies or even intervene remotely, thus reducing the burden on both groups of caregivers [14]. Several studies have shown that telehealth (ie, technological solutions that enable remote monitoring of health status) is effective in improving the well-being and quality of life of CDOAs [15-17]. Technological instruments also allow users to participate actively in their health care and treatment follow-up [18,19]. Despite this potential, the perceptions of CDOAs and their ICs and PCs about the usefulness of technologies in maintaining their health status and helping them to age well at home have only rarely been examined. Previous studies have demonstrated that gerontechnologies are seldom used actively by older adults themselves, and indeed, the vast majority were unaware of their existence [20,21]. However, in recent years, there has been a resurgence of creativity in this field, involving academic researchers, private companies, and public health care services. Questioning the utility of new technologies, their potential uses, and their acceptance and limitations remains relevant [22].

How are technologies that help them remain at home perceived by CDOAs, their ICs, and PCs? More specifically, how do they rate their actual and potential utility?

Objectives

This study, overall, aimed to examine and understand the perceptions and utility of these technologies among physically and cognitively dependent CDOAs, their ICs, and PCs.

Theoretical Framework

This study was guided by the theoretical framework developed by Peek et al [23] and is thus based on the basic components of accepting technology as stated in the results and finding it useful. This framework integrates several factors: (1) utility and ease of use; (2) covering basic needs with technologies; (3) the sense, meaning, acceptance, and value of these technologies; and (4) technologies that help regain or maintain autonomy. According to Peek et al [24], technology can play a role in helping CDOAs remain independent, active, and healthy for as long as possible. Previous studies indicate that the current models of technology acceptance lacked essential predictors, specifically for older adults [25,26].

Methods

Design

This international multicenter study used a qualitative design to collect data on the perceptions of CDOAs, their ICs, and PCs using one-to-one interviews, focus groups (FGs), and photo-elicitation interviews (PEIs). Reporting on the study was based on the checklist for the explicit and comprehensive reporting of qualitative studies [27].

Population and Settings

The study population included CDOA men and women, aged ≥65 years, living in the French department of Haute-Savoie or the French-speaking part of Switzerland's canton Valais, and with a medical prescription or indication for home care. The reason for examining perspectives in 2 countries was not to compare their populations but rather to explore a common issue together (especially in a border area) and codevelop implementable recommendations for clinical practice and technology companies. ICs and PCs involved in clinical practice and employed in 1 of the 5 different community health care centers (2 in France and 3 in Switzerland) were also questioned.

Participant Recruitment

A similar sampling strategy was applied to CDOAs in both countries. On the basis of the comprehensive geriatric assessment database in the Autonomy Scale of Gerontology and Iso-Resource Groups (AGGIR) [28], researchers categorized the functional state of potential participants into 1 of the 3 relevant categories: (1) independent CDOA sustained with meals and cleaning services, (2) mainly physically impaired CDOA, or (3) mainly mild-to-moderately cognitively impaired CDOA.

CDOAs who were unable to understand and speak French or who lacked the capacity for discernment to consent were excluded.

The AGGIR scale includes 2 types of variables: discriminative and illustrative. Discriminative variables evaluate coherence, orientation, washing, dressing, feeding, excretion, transfers, moving indoors, moving outdoors, and remote communication. These elements are assessed via a clinical evaluation carried out in the home of the CDOA by a nurse, who chooses from among 3 criteria for each variable: (1) the person can complete the activity alone, (2) the person can complete the activity partially, or (3) the person cannot complete the activity. Each criterion is also evaluated based on whether the person carries out the activity spontaneously, totally, usually, and correctly.

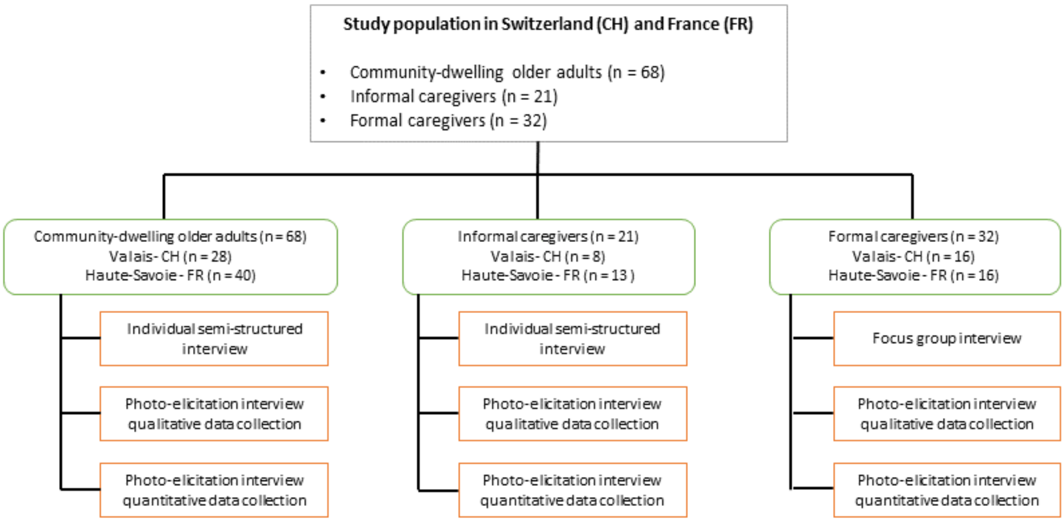
The AGGIR scale is the official national tool in France for comprehensive geriatric assessment under home health care settings. It classifies the geriatric profiles of home-dwelling adults into 6 categories (Groupes Iso-ressources; GIRs). GIRs 5 and 6 designate independent CDOAs who only receive minimal home help, such as meals, cleaning, and shopping. GIRs 3 and 4 designate CDOAs with mostly physical impairments, and GIRs 1 and 2 designate CDOAs with mostly cognitive impairments but also some physical impairments. The AGGIR tool uses clinical observations and questions ICs and older adults themselves on their need for assistance in the essential activities of daily living (ADL). Indeed, AGGIR is based on the topics in the Katz ADL scale, which also classifies CDOAs as independent or physically and/or cognitively impaired. Our study did not use the mini-mental state examination. The research team's geriatrician (NV) clinically judged the participants' physical and mental health statuses. In agreement with the ethical committee's expectations concerning the use of existing clinical data, psychiatric diagnoses were not explored prospectively, as this was beyond the aim of this study. A detailed explanation of the tool is provided in Multimedia Appendix 1.

All 5 centers used randomized sampling to select CDOAs in their respective AGGIR categories (independent, physically impaired CDOAs, and cognitively impaired CDOAs). Multimedia Appendix 1 presents detailed information on categorization using the AGGIR tool. In addition, significant ICs, designated by their CDOAs, were asked to participate in a personal interview and a PEI. The number of ICs recruited was proportional to the CDOA dependency classifications. In collaboration with their supervisors at the 5 community health care centers, purposive samples of PCs involved in direct daily care were invited to participate in 1 of 4 FGs.

Data Collection Procedure

This study was approved by the Human Research Ethics Committee of the Canton Vaud (CER-VD – 2017/000789), Switzerland, and the French Data Protection Authority. Data collection took place from September 2017 to June 2018. After the CDOAs, their ICs, and PCs had given their written informed consent to participate, the study nurse collected the participants' sociodemographic, professional, and health status data (only for the CDOAs) and conducted the scheduled interviews (Figure 1).

Figure 1. Data collection procedure among the community-dwelling older adults, informal caregivers, and formal caregivers.



Community-Dwelling Older Adults

A randomized sample of CDOAs was selected from the 3 categories defined using the AGGIR tool (ie, independent, physically impaired CDOAs, or cognitively impaired CDOAs).

Thus, CDOAs with no limitations according to the AGGIR tool were placed in the independent group, those with physical limitations (eg, traveling and dressing) were categorized in the physically impaired group, and those with cognitive limitations (eg, communication and orientation) joined the cognitively impaired group. Older adults with cognitive and physical limitations were included in the group corresponding to their most pronounced limitation.

Candidates were contacted by telephone and invited to participate. Potential participants received written information about the study before any data were collected. After 48 hours to reflect, a research assistant contacted the potential participants by telephone to address any queries they may have and to obtain their verbal consent to participate in the study. On their acceptance, the assistant made an appointment with the CDOAs for providing more information about the study and to collect the signed consent to participate in the study, before the interview stage. On the day of the interview, the interviewer verified that the participants understood all the information and the implications of participation. Data collection used 10 photographs of relevant technologies and an interview guide developed from a literature review [29]. Audio recordings of the interviews were made.

Informal Caregivers

The CDOAs identified their significant ICs. Theoretical samples of 3 to 4 ICs per type of CDOA (independent, physically impaired, or cognitively impaired) were selected by each of the 5 community health care centers, and they were invited to participate in a personal interview and a PEI. Those without the capacity for discernment were excluded. Audio recordings of the interviews were made.

Professional Caregivers

In collaboration with the supervisors at the 5 community health care centers, 4 FGs were organized (2 in France and 2 in Switzerland), each with 7 to 8 PCs. The research team proposed meeting dates and times for the FGs. The PCs who were willing to participate signed the consent form beforehand, and audio recordings of the FG discussions were made.

Data Collection Instruments

The research team developed and tested guides for the one-to-one interviews and PEIs for the CDOAs and ICs and for the FGs. The interview guides encouraged the participants to discuss relevant research issues by asking open-ended questions. The interviewer could reformulate, reorganize, or clarify questions to gain a deeper understanding of the participant's answers. The guides also helped the participants to better identify practical technologies that might improve health care [30,31]. Data collection using PEIs involved including photographs of relevant technologies into the interview [32,33]. The PEI guide contained 10 photographs, selected using the research team's empirical expertise (Multimedia Appendix 2), from a recent classification of gerontechnologies [29]. The PEI guide was used at the end of the personal interviews and FG discussions. Each participant was asked to select one or more technologies that they thought would be useful to the 3 separate groups of participants (CDOAs, ICs, and PCs). Each photograph was presented in turn with information on the respective technology. The CDOAs were asked, "What do you think of this technology? What could you use it for?" When all the photographs were laid out on the table, they were asked to choose the technology that seemed the most useful or acceptable in their daily life. The ICs were asked, "What do you think about this technology? How could it be useful to you as an informal caregiver? Among all these photographs, could you choose the one that seems the most useful in your daily life?" The PCs were asked, "What do you think when you see this technology? How could you use it? Which of these technologies would you retain, whether for the help it could provide you or for the help it could provide the clients you work with?" Multimedia Appendix 2 presents the technologies included in the PEI.

Data Analysis

Descriptive statistics were compiled to describe the participants' sociodemographic, health, and professional characteristics as well as the technologies selected. Statistical analyses were performed using the IBM Statistical Package for the Social Sciences 25.0 [34]. Personal interviews, PEIs, and FG discussions were transcribed and analyzed using a qualitative content analysis approach [35,36] using NVivo 12 (Qualitative Software International) software [37]. The PEIs were analyzed in parallel using a transversal iterative process and by developing emerging themes. We adopted the approach described by Graneheim and Lundman [35] to ensure the reliability of the results. The interview transcripts were read several times to obtain a sense of the whole. Given the context, units of meaning were condensed into descriptions close to those of the transcripts' contents and, as far as possible, to an interpretation of their underlying meaning (the latent content). Data were condensed for analysis and examination of their content based on the one-to-one interview guide. Once condensed, the units of meaning were considered as a whole and were abstracted into themes. An 8-hour process of reflection and discussion by the research team resulted in an agreed set of themes. The results were presented in the light of the categories taken from the condensed interviews, and the occurrences selected were illustrated with significant examples, including transcriptions from the interviews. For reasons of confidentiality, reporting considered the CDOAs, ICs, and PCs as 3 groups, without distinguishing health care centers or nationalities.

Results

Samples

The study sample was composed of 68 CDOAs, 21 ICs, and 32 PCs. The distribution between the French and Swiss centers was 40 and 28 CDOAs, 13 and 9 ICs, and 16 and 16 PCs, respectively.

Sociodemographic Data

Community-Dwelling Older Adults

Of the 68 CDOAs, almost three-quarters were women (50/68, 74%), with an average age of 82.3 years (SD 7.2; median 82); approximately two-thirds (41/68, 60%) lived in urban areas. Table 1 presents their sociodemographic data and dependency categorization based on the AGGIR scale (Multimedia Appendix 1).

Informal Caregivers

The average age of the 21 ICs was 68.4 years (SD 13.8; median 68), with 16 women and 5 men. Most were retired (n=13), lived in urban settings (n=13), and took care of a loved one with a cognitive impairment (n=12) who required partial assistance with the ADL (n=6; Table 2).

Formal Professional Caregivers

A total of 4 FGs of 8 participants each brought together 32 PCs, all employed at community health care centers. Their average age was 46.7 years (SD 9.4; median 47) and most were female. Their varied professional roles included physicians, nurses, social workers, nursing assistants, care assistants, and occupational therapists (Table 3).

Table 1. The sociodemographic characteristics and functional status of community-dwelling older adults (N=68).

Sociodemographic characteristics and functional status	Values
Age (years)	
Mean (SD)	82.34 (7.2)
Range	64-95
Sex, n (%)	
Male	18 (26)
Female	50 (74)
Place of residence, n (%)	
Rural	27 (40)
Urban	41 (60)
Level of independence, n (%)	
Independent	24 (35)
Physically impaired	23 (34)
Cognitively impaired	21 (31)

Table 2. The sociodemographic characteristics of informal caregivers (N=21).

Sociodemographic characteristics	Value
Age (years)	
Mean (SD)	68.4 (13.4)
Range	43-88
Sex, n (%)	
Male	5 (24)
Female	16 (76)
Loved one’s level of independence, n (%)	
Independent loved one	3 (14)
Physically impaired loved one	6 (29)
Cognitively impaired loved one	12 (57)
Relationship with loved one, n (%)	
Spouse	9 (43)
Daughter/son	9 (43)
Brother	1 (5)
Friend	2 (9)
Occupation, n (%)	
Professionally active	8 (38)
Retired	13 (62)
Place of residence, n (%)	
Rural	8 (38)
Urban	13 (62)

Table 3. The sociodemographic and professional characteristics of professional caregivers (N=32).

Sociodemographic and professional characteristics	Value
Age (years)	
Mean (SD)	46.7 (9.1)
Range	25-60
Sex, n (%)	
Male	3 (9)
Female	29 (91)
Profession, n (%)	
Nurses	11 (34)
Social workers	2 (6)
Occupational therapists	1 (3)
Physicians	3 (10)
Care assistants	10 (31)
Nursing assistants	3 (10)
Supervisors	2 (6)

Findings

We recorded a total of 40 hours and 53 min of personal interviews and PEIs with CDOAs (mean 36 min, SD 13 min; median 35 min; minimum 16 min and maximum 76 min) and 12 hours and 25 min with their ICs (mean 35 min, SD 13 min; median 34 min; minimum 13 and maximum 65 min). We recorded a further 4 hours and 50 min of interviews with the 4 FGs (mean 75 min, SD 6 min; median 72 min; minimum 71 min and maximum 85 min).

In light of the theoretical framework adopted [24], we applied our deductive content analysis along three thematic axes: (1) usefulness and meaning of technology to support the needs of CDOAs, (2) strategies to increase (ease of) technology use among direct and supporting stakeholders, and (3) acceptability of using technology to remain at home among direct and indirect users.

Usefulness and Meaning of Technology to Support the Needs of Community-Dwelling Older Adults

Community-Dwelling Older Adults

CDOAs perceived technology as something that could be useful, especially for other older adults with health problems, but they rarely saw any sense in using them themselves, except when they had significant difficulties with their ADLs. CDOA 6 said that technology was:

> *For people who can't get anything done alone.*

CDOA 63 thought it might be needed in future:

> *Not for the moment, anyway. Maybe in a few years, yes, let's say, if I have problems in a few years, because I have...well, because of my age!"*

CDOA 53 had negative perceptions:

> *Because all these...these modern techniques, and everything, don't interest me much. Because it doesn't match my past life at all.*

CDOA 73, however, held an opposing view:

> *But if something can make life easier, why not? At my age, now's the time to try.*

The concept of technology, in general, was often misunderstood. CDOA 75 exemplified this:

> *Technologies? What do you mean by technology?*

Some CDOAs stated that technology could be useful but not in their current situation. CDOA 9 stated that they did not see any need for it in their current situation:

> *It works automatically, huh? Of course, but for the moment, it's not for me.*

CDOA 13 thought that some devices might be useful to them if they had more severe health problems:

> *I think that I'd have to be really bad, huh, to put this thing on.*

The (positive or negative) interest in technologies was also often stated by this group. CDOA 9's positive perceptions of technology were typical:

> *Well, I think it's...it's 2017; I think it's a necessity, isn't it?*

CDOA 85 spoke out against technology:

> *And then, it's no good at all. Sometimes it's completely harmful, like when things are made available to people, like me, who aren't always able to understand.*

Among CDOAs with mainly cognitive impairments, the word *technology*, without context, was not really understood. As CDOA 82 said:

> *I don't know what that means.*

The relatively *independent* group mainly spoke about how technology could be useful for others. CDOA 7 gave their opinion:

> *Well, it's good for people who don't have many visitors or people who don't get any help from their families. And I'm not one of them. So then, no.*

Misunderstandings about the term and the meaning of *technology* itself, as in the other groups, came up regularly. CDOA 10 expressed this as a lack of knowledge:

> *I don't know which technologies, that's the problem. That's because I don't understand everything...I don't get it...because, well, I know there are a lot of things now, but I don't know everything about them.*

Positive and negative assumptions appeared in similar proportions among the independent group, as they did for the other 2 groups. For example, CDOA 72 gave their opinion on the internet:

> *Today, what I don't agree with is the internet. It's going to destroy the planet. That's it, one way or another because it is so dangerous, especially because of those who know it very well, they could flatten the whole world. So, I'm really not interested in the internet; so I'm not interested in computers either.*

CDOA 83 shared their positive perception:

> *If it's easier for us, if it can help us out, why not, eh?*

Informal Caregivers

Some ICs considered technologies potentially useful in helping CDOAs remain at home, whereas others considered them as indispensable. Their perceptions on the usefulness of a technology seemed to depend mainly on their current state of health, their own openness to technologies, and the costs involved. Different ICs expressed these aspects as follows:

> *The thing that might encourage me is if it, err, her case worsened.* [IC 38]
>
> *She has already been offered them, but she doesn't want them.* [IC 56]
>
> *Yes, change. I think if we change things, I think it would be complicated for her.* [IC 38]
>
> *Oh, well, that must be very expensive.* [IC 43]

Professional Caregivers

PCs perceived technology as useful, but they specified that there was the risk of creating distance in their relationship with the persons in their care and that technology must complete and complement a PC's work, not replace it:

> *It can only be complementary. At some point, it makes sense if a real person has to be involved.* [FG 2]

Moreover, the indication chosen for the use of a specific technology should not only reassure ICs:

> *What worries me is situations like this, where the children have even put cameras in the bedroom, not to monitor their parents, but rather to reassure themselves.* [FG 2]

Technologies that could be introduced to enhance the daily lives of CDOAs include furniture with home automation technologies, suitably adapted public transport, and an elderly friendly urban architecture in environments that facilitate transport mobility:

> *We often have patients who have to take taxis or be driven to hospital or to a medical consultation in town...It's a real transportation problem.* [FG 2]
>
> *Fifteen years ago, they were all at home, they didn't move. Now you see some, I don't know how many, in the city center, using their rollators. It's pretty impressive. Soon, they'll invent rollator parking lots.* [FG 1]

Urban living could also be rethought, perhaps by setting up social centers for CDOAs, on the same principle as youth centers:

> *Architecturally, it would have to be equipped with ramps for the handicapped, for the rooms where they can meet up. As we do for youth centers, why not build a place for older people with physical difficulties?* [FG 1]

PCs highlighted their lack of knowledge about the technological devices available on the market and their lack of training on how to pass on skills to CDOAs and ICs:

> *I don't have much knowledge about this. It's true and it makes sense that some training on this wouldn't be bad.* [FG 2]

Strategies to Increase Ease of Technology Use Among Direct Users and Supporting Stakeholders

Community-Dwelling Older Adults

Encouragement from children and grandchildren, financial help for purchasing technology, and increasing feelings of safety were the main strategies identified by CDOAs themselves for improving the ease of use among their peers. For the groups with mainly *physical* and *cognitive impairments*, encouragement to use technology by *children and grandchildren* was mentioned as an important strategy. Family members often gift devices that they no longer use to their older adult relatives. Children choosing devices, such as remote alarms, was also a facilitator of technology use. CDOAs mentioned that receiving financial aid for the acquisition of a device was a facilitator, as was receiving recommendations from PCs. CDOA 35 exemplified the assistance received from PCs:

> *There is a home for the blind; they proposed it to me. And they sold me the [talking] watch.*

CDOA 76 stated:

> *Yeah, that's right. So, my granddaughter has a tablet, and my son put my number into it.*

Several facilitators to improve beliefs about and attitudes toward technology emerged during the interviews, one being that a device can improve feelings of safety. Financial assistance can be considered a facilitator because it influences people's willingness to acquire a device. Indeed, for the group of *independent* CDOAs, *financial support* for buying a device and receiving help on usage were the main facilitators of their use. CDOA 84 explained their situation:

> *Listen, my son gave me this mobile; I didn't want it. And he said to me, "It'll be easy." Anyway, there you go.*

CDOA 83 said:

> *Someone should come and explain and turn it on and say "You have to do this and do that."*

Again, with regard to beliefs and attitudes, a technology that made participants *feel safer* was the main facilitator of use among this group. CDOA 84 explained this in relation to their personal situation:

> *And then I was unsure, and I was thinking about it, but as I said, I was putting it off and putting it off. And then there's the personal side, you see? I have great neighbors, but the young ones work, and the old ones are no stronger than me, right? So, I said, "Oh, well!" and then it's a bit of a safety thing, even though, as I said, I hope I'll never need it.*

Devices' properties, such as ease of use, also influenced CDOAs. CDOA 10 described using a stair-lift:

> *Yes, usage. You press a button; there's nothing complicated, eh? There are knobs, one at the top, one at the bottom, so it depends on where you are. Well, you grab the knob. And then, if you go up the stairs, you use the stair-lift device.*

Informal Caregivers

Some ICs supported the use of technologies if they found them useful and easy to use. As IC 43 said:

> *If it is a technology that can help me, I will accept it. If it can lighten the burden of helping her.*

Surprisingly, other ICs would only introduce certain technologies if the CDOA's state of health deteriorated. Nevertheless, ICs often felt that they lacked the knowledge to use technologies properly. As IC 71 said:

> *I like the technological side, but I just have one problem with it. It's clear that people who weren't born in the technological age, or who haven't got themselves up to date by 50 or thereabouts (are going to have trouble). I even find that I'm often overwhelmed.*

ICs also noted their relatives' lack of motivation about the introduction of new technologies:

> *Ah, well, it's certainly a very good thing, but she still has to want to do it. She's not on board with all that's modern. So, she'll say no to that* [IC 43]

The ease of use of technologies was overrided, as perceived by the ICs, by the fear of making mistakes when using them or of infringing on the CDOA's privacy and autonomy.

Professional Caregivers

PCs mentioned that the context of technology implementation was important in its ease of use:

> *Older adults just need to be interested in new technologies. They are pushed towards it by their children—by their grandchildren, especially.* [FG 2]

> *Sometimes a neighbor will say, "You know, I've got a rollator, and it goes great," and that's it. It's not some young woman who comes in and says, "This is what you need."* [FG 1]

Indeed, technologies were accepted more readily if relatives and ICs were involved:

> *Sometimes they actually need their family's support in order to be reassured.* [FG 1]

Ease of use also seemed important:

> *There are some things that could improve autonomy as long as they remain simple and easy for people to use.* [FG 2]

According to PCs, the CDOAs who had either found a technology useful or had used technology in their working lives were more willing to use the proposed technologies:

> *Some CDOAs are familiar with Skype, for example, using it with grandchildren abroad or even in another city.* [FG 3]

> *The elderly population that needs these aids now, is often a population that is unaware of the digital revolution.* [FG 4]

Although the point of view of individuals and the context in which technologies are implemented tend to be the factors that promote technology use among CDOAs, the cost involved in acquiring different technologies and the different steps required to obtain them can be seen as barriers:

> *And I find that they're always scared that "It's going to be expensive; I can't." Sometimes they don't realize the means they still have, whether financial or not.* [FG 4]

The emotions associated with technology use can also be a barrier when they generate stress and fear:

> *The more things or equipment involved, the greater the source of stress, for the patient and the family.* [FG 4]

> *You get to a certain point in the aging process where people are not necessarily very accepting of their loss of autonomy. And consequently, in accepting any assistance that might be provided to them.* [FG 4]

Acceptability of Using Technology to Remain at Home Among Direct and Indirect Users

Community-Dwelling Older Adults

For all groups of CDOAs, the acceptability of using technologies was influenced by the self-developed strategy for achieving this purpose, difficulties in using the device, and the technology's cost. As an example of avoiding technology use, several CDOAs said that a simple bedside lamp gave out sufficient light to get to the toilet. The CDOAs found no reason, for example, to replace the home helps who came to clean their home with a robot vacuum cleaner. Help provided directly by humans was the preferred option for many CDOAs. CDOA 13's thoughts on electronic pill dispensers began with a deep breath:

> *I'm telling you, I don't feel the need for one. Oh! If I have something special, my daughter still comes, three times a week, and she is a nurse.*

The CDOAs with mainly *physical impairments* did not find some device properties to be user-friendly, and multiple participants considered this to be the main obstacle. CDOA 35 complained about their tablet:

> *Because handling this device is very complicated for me, you see? I have to turn it all over the place to find out what is going on...No, I'm just not interested.*

Another barrier was the cost of acquiring a new technology, as explained by CDOA 6:

> *Oh, yes, because I can't say I'm going to buy anything, in addition to what we have to buy every day, because we can't afford it.*

For the CDOA group with mainly *cognitive impairments*, the main barrier was the difficulty in using technologies. This was reflected by CDOA 20:

> *So, to start with, I don't want a computer. That's out of the question; I don't want a computer. Well because I wouldn't know how to use it. That's why.*

In several interviews, the group members were found to prefer human contact over technological assistance. CDOA 76 expressed doubts about robots:

Because there's a person, the nurse, working; a robot isn't human, whereas a nurse can talk, say something nice, give you a little smile. But a robot's just "Crash! Bang!"'

Financial hardship was also consistently mentioned as a barrier to technology use. CDOA 76 said:

> *Yeah, robots. Yeah, but you need the means to buy one.*

For the *independent* CDOA group, financial and usage difficulties were also the most frequently mentioned obstacles. Regarding financial aspects, CDOA 10 said:

> *Well, let's say, it's a good thing. It's good, but it must have a cost. Well, yes because...I still have to pay for the nursing home; don't forget that. My husband's pension, unfortunately, only pays for half of it.*

Regarding usage difficulties, CDOA 10 described their tablet:

> *Right now, I do not know how to use it. My daughter, she's got one. My grandsons have one each, but so far I'm not interested.*

Technologies designed to promote older adults' home support, such as fall detectors, GPS, and brain training, are the ones that are most frequently accepted by CDOAs. CDOA 73 explained their acceptance:

> *But we don't know if I'll ever be able to...it's because I'm a little scared, that's why I don't go out alone. I'm a little scared to cross the street. So that's why it's really the fall detector that would suit me best.*

Regarding brain training, CDOA 63 explained:

> *Well, because by...improving your powers of concentration, developing your vocabulary, anyway because of the propositions that are there, I think it's good. That's why.*

Informal Caregivers

Perceptions about the acceptability of technology use were significantly influenced by safety, ease of use, keeping CDOAs autonomous, and the costs of devices.

ICs explained how the acceptability of technologies was related to the safety of their loved ones. This result can be explained by the fact that of the 21 ICs interviewed, 12 were relatives of CDOAs with mainly cognitive impairments. Thus, the technologies chosen by the ICs were very similar to those chosen by the group of mainly cognitively impaired older adults. Opinions of the ICs regarding the acceptability of the technologies presented during the PEI interviews are as follows. For the light path selected by several ICs, the caregivers said:

> *Yes, I think that is really good. Well because my mum gets up at night, she can't find the light.* [IC 42]
>
> *That's right, because I see he gets lost going to the toilet.* [IC 44]

Concerning the fall detector, accepted by about one-third of the ICs, they said:

> *Yes, that could be the most useful there. [...] in the future there could be others [devices].* [IC 43]
>
> *Maybe rather the fall detector, I'd say.* [IC 55]

The ICs were more reluctant about the electronic pill dispenser:

> *Ah, well [...], it could be useful because I want him to have a pillbox; the nurses come to put the pills in. But if he doesn't think about it—if I don't think about it, he doesn't think about it—then it's not much use.* [IC 29]
>
> *Yeah, it's not bad, but I think it's pretty expensive for how much we use it. As long as we're here, well, they're taken at mealtimes, so I'm here [laughs].* [IC 40]
>
> *[…] In my mother's case [...] she'd be unable to take her medication alone.* [IC 42]

The ICs had diverse opinions about robot vacuum cleaners:

> *For me, these robots are good. You shouldn't have anything on the ground [laughs] [...] You shouldn't; it doesn't go in corners, it's round [laughs].* [IC 29]
>
> *Well, it depends on how much it costs, huh?* [IC 38]
>
> *Not essential, in my opinion.* [IC 40]
>
> *[….] I think not. No, it might be more dangerous than useful.* [IC 42]

As for the opinions on the service robot, the ICs said:

> *Yes, listen, yes, it's good, but in the end, a human being is still better.* [IC 23]
>
> *[…] I find it rather stressful.* [IC 29]

Opinions on the usefulness of a GPS bracelet could be summed up by this statement:

> *That, I, I'm for that. If she ever decided to go somewhere, it would let her know where she was...it would let us know where she was.* [IC 38]

Only a fraction of the ICs found the touchpad to be an interesting technology. There were different opinions on this:

> *No, so my mom, my mom... she's not interested.* [IC 23]
>
> *It might make things easier for him, but since he was not very enthusiastic, he quickly lost patience.* [IC 29]
>
> *Uh, no [laughs], there. Uh, I think it's impossible. Because even a mobile phone would be complicated to use, so...He still manages to work the TV, but that's all.* [IC 55]

Among the ICs interviewed, the idea of a social network for ICs raised little interest:

> *I think it's good… For the moment, I don't need it, because...Well, I have a good rapport with the home help; it's going well.* [IC 23]
>
> *I'm not…I've never been a person like that, not very communicative and that...So, everyone has to say what they do, what they've got...It's not...No, not for me. But I'm not saying that it might not be a help to a lot of people.* [IC 29]
>
> *Oh! It's not bad. It could...for someone who is alone, be a help, yes.* [IC 40]

Brain training technology was selected by only one IC, but the quoted text below illustrates the difficulties in using this technology:

> *She can't do it alone either. Well, memory exercises would be great, but maybe someone would have to help because of her not being able to use a tablet.* [IC 42]

To the ICs, activity sensors were not a promising technology to strengthen home support:

> *Ah! Well, for people who live alone, that's very good.* [IC 23]
>
> *Oh, yeah, sometimes he hasn't closed [the fridge door] properly. It's happened sometimes, when it didn't shut properly.* [IC 29]
>
> *[…] she'll say to me: "You don't think I'm crazy yet, do you?" All this scares me.* [IC 38]
>
> *Yes. Yes, well that could be, err, that could be good, indeed. Yeah because in half an hour anything can happen, eh?* [IC 42]
>
> *[…] I wouldn't even trust the sensor [...].* [IC 44]
>
> *[…] it's true that these are technologies that are great for those who are not safe, but who are in their right mind, and everything, who don't have all their heads and bodies; and who may be putting themselves in danger, uh, about that.* [IC 55]

Professional Caregivers

Among PCs, perceptions about the acceptability of technology differed according to the constraints that technologies impose on the CDOAs. The more technologies focus on monitoring a CDOA's behavior, the more ethical questions they raise. On the contrary, the more they increase a CDOA's autonomy and safety and decrease the burden on ICs, the more the PCs promoted them. Opinions on the acceptability of some of the technologies proposed during the PEIs were as follows. The light path technology was broadly well-received by the PCs, and the following statement illustrates their favorable opinion of this technology:

> *A technology that should be introduced and available for all older people.* [FG 1]

Technologies that increase safety, such as the fall detector, were mostly preferred by the PCs:

> *[…] if it's good, it can be very useful.* [FG 2]

The PCs showed little interest in the robot vacuum cleaner and the service robot:

> *[…] There's a risk of falling over the vacuum cleaner when it's running...* [FG 4]
>
> *[…] It's interesting if it assists caregivers [gives them time to do another activity while the robot is working].* [FG 1]
>
> *There's a risk of replacing humans […].* [FG 3]

Several PCs had questions regarding the GPS bracelet technology:

> *[…] it's not relevant if its only purpose is to calm the family […]* . [FG 3]
>
> *[…] there are ethical problems if it's used as a means to monitor the person […].* [FG 3]

The touch tablet and brain training technologies were rarely chosen by the PCs. Their reluctance was expressed as follows:

> *You have to accompany the arrival of the tablet in the person's environment and ensure that the person is interested in using it….* [FG 4]
>
> *It might be interesting to couple a tablet to the remote alarm [to enable visualizing the person's condition if they can't or can no longer express themselves because of cognitive disorders]...* [FG 4]
>
> *[…] difficult to offer if the person does not know how to use a tablet…* [FG 4]
>
> *...can be a prevention tool…* [FG 3]

Although some PCs selected the social network and activity sensor technologies, there were mixed opinions:

> *[Social networking] technology is not necessarily appropriate if the caregiver does not understand those technologies; it's rather aimed at young caregivers [...].* [FG 2]
>
> *[…] it can make caregivers feel supported [...].* [FG 2]

Concerning activity sensors, this response was obtained:

> *[...] Pose an ethical dilemma [...]; [...] culture can be an obstacle to the use of this technology; [...] interesting technology, and install the sensors only as needed [to see if the fridge is opening in case of a heatwave; to see that people are getting out of bed]. It can be a complementary tool in food monitoring if this problem has been previously identified [...]* [FG 2]

When choosing the most useful technology overall, the PCs differed according to their profession. Physicians, nurses, and occupational therapists chose the fall detector, whereas nursing assistants, social workers, and direct community health providers chose social network. Table 4 presents a summary of the main topics covered.

Multimedia Appendix 3 provides detailed data on the technologies preferred by CDOAs, ICs, and PCs. Multimedia Appendix 4 presents a detailed overview of the perceived barriers to technology use among CDOAs, as well as the photographs selected during their PEIs. Multimedia Appendix 5 provides a detailed overview of CDOAs' perceptions of technology and its facilitators.

Table 4. Themes and subthemes highlighted by the thematic content analysis.

Themes	Subthemes		
	CDOAs[a]	ICs[b]	PCs[c]
Usefulness and meaning of technology to support the needs of CDOAs	• Useful for others in my peer group • Useful later in my life • Lack of meaning • Technology often misunderstood	• Useful depending on the CDOA's state of health • The CDOA's openness to using technologies • Technology costs	• Useful to sustain the ADL[d] • Risk of creating distance in relationships with patients • Fear of being replaced by technology
Strategies to increase ease of technology use among direct users and supporting stakeholders	• Encouragement from children/grandchildren • Financial support • Increase feeling of safety	• Easing the daily strain for ICs and PCs • Lightening the burden of helping • Fear of mistakes	• Appropriate technology implementation strategy • Involving relatives • Increasing the autonomy of CDOAs
Acceptability of using technology to remain at home among direct and indirect users	• Influenced by having a self-developed strategy to achieve purposes • Difficulties using devices • The cost of devices • Increasing the ability to remain at home	• Safety first • Ease of use • Staying autonomous • The cost of devices	• Constraints that technologies impose • Ethical questions if the focus is on monitoring behavior • Increasing the CDOA's autonomy and safety • Lessing the burden on ICs

[a] CDOA: community-dwelling older adults.

[b] IC: informal caregiver.

[c] PC: professional caregiver.

[d] ADL: activities of daily living.

Discussion

Principal Findings and Comparison With Previous Work

To the best of our knowledge, this is the first research in the field to have combined personal interviews, FGs, and PEIs in the same study. Our results show that technologies are not yet well integrated into the daily care of frail or cognitively impaired CDOAs. Although technology can certainly offer solutions to some of the specific impairments typical among CDOAs, ICs and PCs still have little knowledge about their existence and how they should use them in everyday practice. The heterogeneity of participants' profiles made it difficult to compare this study with others involving more homogeneous samples [38,39].

Our results indicate that most participants had both positive and negative attitudes toward technology. Positive attitudes were most often related to technologies that ensured continued mobility at a lower cost and with advanced functions (a GPS bracelet). Negative attitudes of CDOAs were often associated with the risks of becoming technology–dependent and the risk of fewer or poorer interactions with ICs and PCs, the realization that they were no longer able to stay at home, potential social problems (disturbing ICs and PCs with alarms), and the financial costs of technology use. Furthermore, the findings suggest that ICs had more positive attitudes toward assisting technology and an increased acceptance of technology than the CDOAs themselves. This could be explained by the fact that the average participating IC was 14 years younger than their CDOA and thus had more contact with technology in their daily life [40]. Nevertheless, some ICs evoked their feelings of being overwhelmed by technology. An analysis of their sociodemographic profiles shows that these ICs were generally older relatives, which is in line with the hypothesis that younger ICs are more open to using technologies. Another possible explanation is related to the feeling of being burdened, which urges ICs to seek technological support to alleviate their problems [41].

Technology use was often conditioned by the multiple expected outcomes CDOAs had with regard to remaining at home, risk control by ICs and PCs, and trying to avoid hospitalization or institutionalization. In contrast, the nonuse of technology was linked to the participants' use of their personal and individual capacities, their health, their physical and intellectual functional capacities, and the environmental barriers they encountered (eg, risk of falling). To encourage the participants to adopt technology, our findings suggest that the potential barriers to this should first be removed at the individual level (each situation is unique, including costs), then at the technological level (technological flexibility adaptable to the situation), and finally at the environmental level (life context). These results were consistent with the research conducted by Chen and Chan [30].

Results from the PC FGs revealed that they had mixed attitudes toward technology, although most were interested in specific technologies for everyday use. They highlighted that each care situation was unique, uncertain, and often complex. They were particularly interested in the development of new technologies to help with the ADL. Finally, barriers to the use of technology by PCs themselves included lack of interest, a need for training, the architectural inadequacies of house design, cost issues, and their fears of being replaced by that technology and of the risks of dehumanizing care relationships with CDOAs. Our results highlight the need to pay more attention to adapting technology to users' personal preferences, while focusing on technologies that provide solutions to individual problems. However, to date, there is little evidence to evaluate the usefulness or effectiveness of certain technologies in home care and notably to avoid hospitalization [42]. Causal intervention research expects a coconstruction between users, inventors, researchers, and manufacturers. This would make it possible to develop technologies that meet users' needs (eg, a device that detects falls and sends an alert) and to fulfill the commercial requirements for making those technologies accessible to large numbers of CDOAs [43]. Our findings suggest that most devices will not require significant technological breakthroughs but rather careful adaptations to the specificities of end users. Many technologies will still require an assessment of user needs and their real potential for use. Technology development must consider the somatic and psychopathological states of health of CDOAs, as mentioned in Cohen-Mansfield's work [33]. Our findings suggest that it is important to go beyond the myth that technology will replace ICs and PCs. Indeed, integrating technologies into the home care activities of PCs should be promoted, as this would give them the time to do other things [44,45]. Technologies for care practices are promoted as facilitating safety and independent living as well as avoiding or delaying institutionalization. However, there is still a gap between these goals and the complex realities in which technologies are used [46].

Technologies in home health care settings must also be adapted to the needs and concerns of PCs [44]. In line with previous research, complex surveillance technologies (eg, cameras and motion detectors) were perceived to be intrusive and as posing a high risk to privacy; they were widely rejected by home health care stakeholders [47]. PCs raised concerns that the data collected might be stored for a long time and be accessible and misused by other actors (eg, insurance companies) [13]. The complex context surrounding care for frail CDOAs is unique, uncertain, and constantly changing [48]. This study showed that ICs were often involved in the daily care of a loved one, but that they rarely used technologies. Slowly introducing effective technologies into the everyday care they give would be a way to make their support more effective, either for monitoring care or in maintaining their relationship with the CDOA. This would help to maintain the health of both CDOAs and their ICs, who often become overwhelmed by their care responsibilities, leading to stress and physical and mental exhaustion.

The development of technologies to strengthen home support and prevent the loss of autonomy is a demographic, human, social, and economic imperative; it must be accompanied by the development of multidisciplinary skills [49]. Academic institutions should supervise this development, particularly by proposing critical reflection. It should also be noted that the realities of life will force both researchers in innovative technologies and prescribers to accept that the effectiveness of a technological tool should be based primarily on pragmatic approaches that lack the scientific rigor of causality [50].

Strengths and Limitations

The significant sample of CDOAs, ICs, and PCs involved in this study suggests that our results may be transferable to other regions of Western Europe. The PEIs made it possible to clarify perceptions and specify acceptable technologies. Another strength of this study was its use of a novel combination of methodological approaches, via one-to-one interviews, FGs, and PEIs. This approach illustrated which technologies might be acceptable and useful to CDOAs, ICs, and PCs in the context of home care support. Our research population's heterogeneity did not permit a thorough comparison with other studies in the literature as they involved more homogeneous samples [18-20]. In addition, to the best of our knowledge, this is the first paper in the field to combine one-to-one interviews, FGs, and PEIs, which enriches the value of our results but complicates their comparison with previous studies. One limitation of this study concerns the AGGIR tool's subjectivity when used as an overall clinical classification system of impairment among CDOAs. In contrast, other tools are more oriented toward disease and physiopathological classification (Resident Assessment Instrument-Home care). Finally, this study had some methodological limitations concerning the choices of the images shown in the PEIs. These could not always best present the different gerontechnologies to the heterogeneous research participants (CDOAs, ICs, and PCs).

Conclusions

Despite the omnipresence of technology in modern society, research into perceptions about it and its use to improve the daily lives of frail and cognitively impaired CDOAs is poor compared with research into its impact on younger populations. Despite some evidence that technologies promote independence among CDOAs, they are often underused in daily home health care. With this in mind, this study explored the perceptions about technology's place in home health care among CDOAs, ICs, and PCs. Our findings showed that although many technologies were available to support independent living for CDOAs, they were significantly underused. Our results also affirmed that efforts should be made to adapt technologies to the needs of CDOAs, their ICs, and PCs. It seems important to continue searching for empirical evidence of the relevance and effectiveness of new technologies. This will help to specify the most important areas requiring intervention and indications for technology use among different profiles of CDOAs, their ICs, and PCs. Technology will also help to optimize the management of CDOAs' health problems and slow their loss of autonomy, both of which will strengthen home health care. Progress toward this goal will only be achieved through close cooperation between technical experts, home health care experts, and the end users—CDOAs, ICs, and PCs—who need the appropriate technological tools to meet their needs and expectations. Finally, as the mean age of the population is rising, the proportion of older adults with an interest in technology and with technological skills will increase. The findings in this study will enable future CDOAs to clearly express the advantages and limitations of the technologies in their lives. The current situation is therefore very fluid, and research will have to adjust to this dynamic process. Although there is no doubt that technologies will play an increasingly important role in health care services for CDOAs and in the work of their ICs and PCs, it is more difficult to predict which types of interventions may develop. Establishing convincing results based on robust scientific evidence will be difficult, but clinical research will play a key role.

Acknowledgments

The authors wish to thank Christelle Mercier, Françoise Martinet, and Dominique German for their contributions to data collection and their suggestions during the writing of the manuscript. This work was supported by the European cross-border cooperation program Interregion France-Switzerland V 2014-2020, whose total budget is EUR 1,394,247.16 (US $1,401,080.54). The French side of the project received EUR 701,895.75 (US $705,336.42) from the European Regional Development Fund. The Swiss side of the project was supported by CHF 187,503.34 (US $183,826.80) of federal funds from INTERREG Switzerland and by CHF 75,000 (US $73,529.41) and CHF 50,000 (US $49,019.61) of cantonal funds from Vaud and Valais, respectively.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Classification of community-dwelling older adults degree of dependence: the Autonomy Scale of Gerontology and Iso-Resource Groups model.

[DOCX File , 17 KB - jmir_v22i6e17930_app1.docx]

Multimedia Appendix 2

Interview guides.

[DOCX File , 1865 KB - jmir_v22i6e17930_app2.docx]

Multimedia Appendix 3

Preferred technology of informal caregivers and professional caregivers.

[DOCX File , 17 KB - jmir_v22i6e17930_app3.docx]

Multimedia Appendix 4

Technology selected by community-dwelling older adults (n=68).

[DOCX File , 15 KB - jmir_v22i6e17930_app4.docx]

Multimedia Appendix 5

Perceptions of technology and facilitators of technology use among community-dwelling older adults (N=68).

[DOCX File , 14 KB - jmir_v22i6e17930_app5.docx]

References

1. OECD Factbook 2015-2016: Economic, Environmental and Social Statistics. Paris: OECD Publishing; 2016.
2. Boldy D, Grenade L, Lewin G, Karol E, Burton E. Older people's decisions regarding 'ageing in place': a Western Australian case study. Australas J Ageing 2011 Sep;30(3):136-142. [doi: 10.1111/j.1741-6612.2010.00469.x] [Medline: 21923707]
3. Goldman DP, Chen C, Zissimopoulos J, Rowe JW, Research Network on an Aging Society. Opinion: measuring how countries adapt to societal aging. Proc Natl Acad Sci USA 2018 Jan 16;115(3):435-437 [FREE Full text] [doi: 10.1073/pnas.1720899115] [Medline: 29339547]

4. North MS, Fiske ST. Modern attitudes toward older adults in the aging world: a cross-cultural meta-analysis. Psychol Bull 2015 Sep;141(5):993-1021. [doi: 10.1037/a0039469] [Medline: 26191955]
5. Pruchno RA, Smyer MA. Challenges of an Aging Society. Boston: Johns Hopkins University Press; 2007.
6. OECD. Old-age dependency ratio. In: Pensions at a Glance 2017: OECD and G20 Indicators. Paris: OECD; 2017.
7. Bouma H. Gerontechnology: emerging technologies and their impact on aging in society. Stud Health Technol Inform 1998;48:93-104. [Medline: 10186579]
8. Franco A, Rialle V. Loss of autonomy and home support. Article in French. Perte d'autonomie et maintien à domicile. Esprit 2010;7(7):85-92. [doi: 10.3917/espri.1007.0085]
9. Fozard JL, Rietsema J, Bouma H, Graafmans JA. Gerontechnology: creating enabling environments for the challenges and opportunities of aging. Educ Gerontol 2000 Jun;26(4):331-344. [doi: 10.1080/036012700407820]
10. Margolis E, Pauwels L. Some theortical and methodological views on Photo-Elicitation. In: The SAGE Handbook of Visual Research Methods. Thousand Oaks, CA: Sage Publications; 2011:201-213.
11. Cornet G, Carré M. Technologies for the care, autonomy and social bond of the elderly: what's new?. Article in French. Technologies pour le soin, l'autonomie et le lien social des personnes âgées : quoi de neuf? Gérontologie et Société 2008;31(126):113-128. [doi: 10.3917/gs.126.0113]
12. Bronswijk J, Bouwhuis DG, Fozard JL, Bouma H. Gerontechnology's basics. Gerontechnology 2008;7(2):80. [doi: 10.4017/gt.2008.07.02.017.00]
13. Schulz R, Wahl H, Matthews JT, De Vito Dabbs A, Beach SR, Czaja SJ. Advancing the aging and technology agenda in Gerontology. Gerontologist 2015 Oct;55(5):724-734 [FREE Full text] [doi: 10.1093/geront/gnu071] [Medline: 25165042]
14. Woolrych R. Gerontechnology: creating enabling environments to meet the challenges and opportunities of an aging society. Medicina 2016;49(Suppl 2):5-6 [FREE Full text]
15. Slegers K, van Boxtel MP, Jolles J. Effects of computer training and internet usage on the well-being and quality of life of older adults: a randomized, controlled study. J Gerontol B Psychol Sci Soc Sci 2008 May;63(3):P176-P184. [doi: 10.1093/geronb/63.3.p176] [Medline: 18559683]
16. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of social robots in mental health and well-being research: systematic review. J Med Internet Res 2019 Jul 24;21(7):e13322 [FREE Full text] [doi: 10.2196/13322] [Medline: 31342908]
17. Handel MJ. mHealth (mobile health)-using Apps for health and wellness. Explore (NY) 2011;7(4):256-261. [doi: 10.1016/j.explore.2011.04.011] [Medline: 21724160]
18. Vizer LM, Eschler J, Koo BM, Ralston J, Pratt W, Munson S. 'It's not just technology, it's people': constructing a conceptual model of shared health informatics for tracking in chronic illness management. J Med Internet Res 2019 Apr 29;21(4):e10830 [FREE Full text] [doi: 10.2196/10830] [Medline: 31033452]
19. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, van Houten H, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Arch Intern Med 2012 May 28;172(10):773-779 [FREE Full text] [doi: 10.1001/archinternmed.2012.256] [Medline: 22507696]
20. Alemdar H, Ersoy C. Wireless sensor networks for healthcare: a survey. Comput Netw 2010 Oct;54(15):2688-2710. [doi: 10.1016/j.comnet.2010.05.003]
21. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020 Feb;39(2):141-150. [doi: 10.1177/0733464818807468] [Medline: 30353776]
22. Yusif S, Soar J, Hafeez-Baig A. Older people, assistive technologies, and the barriers to adoption: a systematic review. Int J Med Inform 2016 Oct;94:112-116. [doi: 10.1016/j.ijmedinf.2016.07.004] [Medline: 27573318]
23. Peek ST, Wouters EJ, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJ. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
24. Peek ST, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
25. Marziali E. Gerontechnology: research and practice in technology and aging. Canadian Journal on Aging 2006;25(2):237-238 [FREE Full text]
26. Marschollek M, Künemund H. [Gerontechnology between acceptance and evidence: results of the Lower Saxony Research Network 'Design of Environments for the Ageing']. Z Gerontol Geriatr 2014 Dec;47(8):639-640. [doi: 10.1007/s00391-014-0828-1] [Medline: 25348949]
27. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
28. Benaim C, Froger J, Compan B, Pélissier J. [The assessment of autonomy in elderly people]. Ann Readapt Med Phys 2005 Jul;48(6):336-340. [doi: 10.1016/j.annrmp.2005.04.005] [Medline: 15932780]
29. Fernández-Caballero A, González P, Navarro E. Gerontechnologies - Current achievements and future trends. Expert Syst 2017 Apr 10;34(2):e12203. [doi: 10.1111/exsy.12203]
30. Chen K, Chan A. Use or non-use of gerontechnology--a qualitative study. Int J Environ Res Public Health 2013 Sep 30;10(10):4645-4666 [FREE Full text] [doi: 10.3390/ijerph10104645] [Medline: 24084674]

31. Ben-Ahmed L. The question of the stigmatization of gerontechnologies of remote assistance and geolocation. Article in French. La question de la stigmatisation des gérontechnologies de téléassistance et géolocalisation. Les cahiers de l'année gérontologique 2012;4(4):394-397 [FREE Full text]
32. Bates EA, McCann JJ, Kaye LK, Taylor JC. 'Beyond words': a researcher's guide to using photo elicitation in psychology. Qual Res Psychol 2017 Jul 25;14(4):459-481. [doi: 10.1080/14780887.2017.1359352]
33. Cohen-Mansfield J, Biddison J. The scope and future trends of gerontechnology: consumers' opinions and literature survey. J Technol Hum Serv 2007;25(3):1-19. [doi: 10.1300/j017v25n03_01]
34. IBM. Somer, NY: IBM Corporation; 2016. Statistical Package for Social Sciences 25.0 URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2020-05-27]
35. Graneheim U, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
36. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open 2016;2:8-14. [doi: 10.1016/j.npls.2016.01.001]
37. QSR International. 2015. NVivo - version 12 URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/support-services/nvivo-downloads [accessed 2020-05-27]
38. Whelan S, Murphy K, Barrett E, Krusche C, Santorelli A, Casey D. Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. Int J of Soc Robotics 2018;10(5):643-668. [doi: 10.1007/s12369-018-0471-x]
39. Steventon A, Bardsley M, Billings J, Dixon J, Doll HM, Beynon M, et al. Effect of telecare on use of health and social care services: findings from the Whole Systems Demonstrator cluster randomised trial. Age Ageing 2013 Jul;42(4):501-508 [FREE Full text] [doi: 10.1093/ageing/aft008] [Medline: 23443509]
40. Krick T, Huter K, Domhoff D, Schmidt A, Rothgang H, Wolf-Ostermann K. Digital technology and nursing care: a scoping review on acceptance, effectiveness and efficiency studies of informal and formal care technologies. BMC Health Serv Res 2019 Jun 20;19(1):400 [FREE Full text] [doi: 10.1186/s12913-019-4238-3] [Medline: 31221133]
41. Marasinghe KM. Assistive technologies in reducing caregiver burden among informal caregivers of older adults: a systematic review. Disabil Rehabil Assist Technol 2016;11(5):353-360. [doi: 10.3109/17483107.2015.1087061] [Medline: 26371519]
42. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
43. Jansen S, Boye N, Becker C, Mellone S, Chiari L. Fall prevention and gerontechnology. Eur Geriatr Med 2013 Sep;4:S2. [doi: 10.1016/j.eurger.2013.07.054]
44. Ma X, Liu Y, Roberto KA, Reed J. Using wireless sensing technology for medication management by older adults. Gerontologist 2015;55(Suppl_2):390. [doi: 10.1109/SENSORCOMM.2009.32]
45. Cook DJ. Health monitoring and assistance to support aging in place. J Univers Comput Sci 2006;12(1):15-29. [doi: 10.3217/jucs-012-01-0015]
46. Dantoine T, Daviet J, Nibeaudeau P, Porte H, Fabre S, Huon J. editors. Solutions innovantes de demain pour l'assistance à domicile. Silver-Economie 2015 [FREE Full text]
47. Offermann-van Heek J, Ziefle M. They don't care about us! Care personnel's perspectives on ambient assisted living technology usage: scenario-based survey study. JMIR Rehabil Assist Technol 2018 Sep 24;5(2):e10424 [FREE Full text] [doi: 10.2196/10424] [Medline: 30249592]
48. Sixsmith A. Vieillesse et technologie : du laboratoire au monde réel. Angewandte GERONTOLOGIE Appliquée 2016 Jan;1:26-27. [doi: 10.1024/2297-5160/a000015]
49. Peine A, Neven L. From intervention to co-constitution: new directions in theorizing about aging and technology. Gerontologist 2019 Jan 9;59(1):15-21. [doi: 10.1093/geront/gny050] [Medline: 29850812]
50. Peine A. Technology and ageing—theoretical propositions from Science and technology studies (STS). In: Neves BB, Vetere F, editors. Ageing and Digital Technology: Designing and Evaluating Emerging Technologies for Older Adults. Singapore: Springer Singapore; 2019:51-64.

Abbreviations

ADL: activities of daily living
AGGIR: Autonomy Scale of Gerontology and Iso-Resource Groups
CDOA: community-dwelling older adult
FG: focus group
GIR: Groupes Iso-ressources
IC: informal caregiver
PC: professional caregiver
PEI: photo-elicitation interview

Edited by G Eysenbach; submitted 22.01.20; peer-reviewed by K Jekel, S Fritz, N Heynsbergh; comments to author 12.02.20; revised version received 28.04.20; accepted 30.04.20; published 04.06.20.

Please cite as:
Verloo H, Kampel T, Vidal N, Pereira F
Perceptions About Technologies That Help Community-Dwelling Older Adults Remain at Home: Qualitative Study
J Med Internet Res 2020;22(6):e17930
URL: http://www.jmir.org/2020/6/e17930/
doi:10.2196/17930
PMID:32496197

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org